ANL/ESD-14-19

Dynamic Factor Graphs – A New Wind Power Forecasting Approach

Energy Systems Division

About Argonne National Laboratory
Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

DOCUMENT AVAILABILITY

Online Access: U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via DOE's SciTech Connect (http://www.osti.gov/scitech/).

Reports not in digital format may be purchased by the public from the National Technical Information Service (NTIS):
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312
www.ntis.gov
Phone: (800) 553-NTIS (6847) or (703) 605-6000
Fax: (703) 605-6900
Email: **orders@ntis.gov**

Reports not in digital format are available to DOE and DOE contractors from the Office of Scientific and Technical Information (OSTI):
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
www.osti.gov
Phone: (865) 576-8401
Fax: (865) 576-5728
Email: **reports@osti.gov**

Disclaimer
This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.

ANL/ESD-14-19

DYNAMIC FACTOR GRAPHS – A NEW WIND POWER FORECASTING APPROACH

by

Cátia Silva
Ricardo Bessa
Erika Pequeno
Jean Sumaili
Vladimiro Miranda

INESC Technology and Science (INESC TEC)

and

Zhi Zhou
Audun Botterud

Argonne National Laboratory

September 2014

This page is intentionally blank.

CONTENTS

ACKNOWLEDGMENTS x

ABBREVIATIONS xi

ABSTRACT 1

1 INTRODUCTION 5

2 TIME SERIES 7

2.1 Time Series Modeling without Hidden Variables 7
2.2 Time Series Modeling with Hidden Variables 7

3 DYNAMIC FACTOR GRAPHS 9

3.1 Formalism 9
3.2 Factors Used 10
3.3 Maximum Likelihood Energy-Based Inference 11
3.3.1 Energy as Negative Log-Likelihood 11
3.3.2 Inference of Latent Variables 12
3.4 Expectation Maximization-Based Learning of DFGs 13
3.4.1 Expectation Maximization Algorithm 13
3.4.2 Simplification and Approximation 13
3.4.3 Alternated E-Step and M-Step Procedure 13
3.5 Features 14
3.6 Performance Measures 15

4 APPLICATION TO TIME SERIES MODELS AND DYNAMIC SYSTEMS 17

4.1 Model Training Approach 19

5 EMPIRICAL RESULTS: MIDWEST WIND FARM A 21

5.1 10-Minute-Frequency Dataset 21
5.2 30-Minute-Frequency Dataset 25
5.3 1-Hour-Frequency Dataset 27
5.4 Midwest Wind Farm A for Another Time Period 30
5.4.1 Including Explanatory Variables 33

6 EMPIRICAL RESULTS: MIDWEST WIND FARM B 35
6.1 10-Minute-Frequency Dataset 35
6.2 30-Minute-Frequency Dataset 37
6.3 1-Hour-Frequency Dataset 40

CONTENTS (CONT.)

7 EMPIRICAL RESULTS: SOUTH CENTRAL WIND FARM C 43

7.1 10-Minute-Frequency Dataset 43
7.2 30-Minute-Frequency Dataset 45
7.3 1-Hour-Frequency Dataset 48

8 WIND TO POWER MODEL WITH NWP RESULTS 51

8.1 Midwest Wind Farm A 51
8.1.1 Input Data 51
8.1.2 Results of Performance Analysis 52
8.1.3 Evaluation of the Wind Speed NWPs 54
8.1.4 Summary of Results 63
8.2 South Central Wind Farm D 63
8.2.1 Input Data 63
8.2.2 Results of Performance Analysis 64
8.2.3 Summary of Results 70

9 CONCLUDING REMARKS 71

9.1 Future Research 72

10 REFERENCES 73

APPENDIX A: MAXIMUM A POSTERIORI APPROXIMATION 75

APPENDIX B: DERIVATION OF THE GRADIENT FUNCTION 77

FIGURES

1 Architecture of a DFG that describes the structure of a general state-space model 9

2 Architecture of a DFG with a more complex structure 9

3 DFG structure for a case in which the time series Y is considered the output of a system 10

4 General description of a factor linking variables a_t and o_t through function g, with energy term $E(a_t, o_t)$ 11

FIGURES (CONT.)

5 WFA for 10-minute- time step: RMSE improvements w.r.t persistence for the DFG models with Markov order 1 and the ARIMA model, with an initial training set with 1 month of observations 22

6 WFA for 10-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and the ARIMA model, with an initial training set with 1 month of observations 23

7 WFA for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 24

8 WFA for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 24

9 WFA for 30-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 25

10 WFA for 30-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 26

11 WFA for 30-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 26

12 WFA for 30-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 27

13 WFA for 60-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 28

14 WFA for 60-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 28

15 WFA for 60-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 29

16 WFA for 60-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 29

17 WFA for 1-hour time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 30

18 WFA for 1-hour time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 31

FIGURES (CONT.)

19 WFA for 1-hour time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 32

20 WFA for 1-hour time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 32

21 WFA for 1-hour time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 using the two components of the wind speed and the wind speed modulus as explanatory variables 33

22 WFA for 1-hour time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 using the two components of the wind speed and the wind speed modulus as explanatory variables 34

23 WFB for 10-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 35

24 WFB for 10-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 36

25 WFB for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 36

26 WFB for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 37

27 WFB for 30-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 38

28 WFB for 30-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 38

29 WFB for 30-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 39

30 WFB for 30-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 39

31 WFB for 60-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 40

32 WFB for 60-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 41

FIGURES (CONT.)

33 WFB for 60-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 41

34 WFB for 60-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 42

35 WFC for 10-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 43

36 WFC for 10-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 44

37 WFC for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 44

38 WFC for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 45

39 WFC for 30-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 46

40 WFC for 30-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 46

41 WFC for 30-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 47

42 WFC for 30-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 47

43 WFC for 60-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 48

44 WFC for 60-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA 49

45 WFC for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 49

46 WFC for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1 50

47 WFA: NMAE results comparison 52

FIGURES (CONT.)

48 WFA: NRMSE results comparison 53

49 WFA: NBIAS results comparison 53

50 WFA: NSDE results comparison 54

51 WFA: Power curve for 30-minute data with predicted and measured wind speeds 55

52 WFA: NMAE comparison between NWP wind speed and persistence forecasts for 30-minute data 56

53 WFA: NMAE improvement over persistence for 30-minute data 56

54 WFA: NRMSE comparison between NWP wind speed and persistence forecasts for 30-minute data 57

55 WFA: NRMSE improvement over persistence for 30-minute data 57

56 WFA: NSDE comparison between NWP wind speed and persistence forecasts for 30-minute data 58

57 WFA: NBIAS comparison between NWP wind speed and persistence forecasts for 30-minute data 58

58 WFA: Power curve for 1-hour data 59

59 WFA: NMAE comparison between NWP wind speed and persistence forecasts for 1-hour data 60

60 WFA: NMAE improvement over persistence for 1-hour data 60

61 WFA: NRMSE comparison between NWP wind speed and persistence forecasts for 1-hour data 61

62 WFA: NRMSE improvement over persistence for 1-hour data 61

63 WFA: NSDE comparison between NWP wind speed and persistence forecasts for 1-hour data 62

64 WFA: NBIAS comparison between NWP wind speed and persistence forecasts for 1-hour data 62

FIGURES (CONT.)

65 WFD: NMAE comparison between the two prediction models and persistence..... 65

66 WFD: NRMSE comparison between the two prediction models and persistence..... 65

67 WFD: NBIAS comparison between the two prediction models and persistence 66

68 WFD: NMAE improvement over persistence for the two prediction models 66

69 WFD: NRMSE improvement over persistence for the two prediction models 67

70 WFD: NMAE comparison between the two prediction models and persistence..... 68

71 WFD: NRMSE comparison between the two prediction models and persistence..... 68

72 WFD: NBIAS comparison between the two prediction models and persistence 69

73 WFD: NMAE improvement over persistence for the two prediction models 69

74 WFD: NRMSE improvement over persistence for the two prediction models 70

TABLE

1 Parameters and their values 18

ACKNOWLEDGMENTS

This document has been produced by the Centre for Power and Energy Systems of INESC TEC and Argonne National Laboratory. The INESC TEC team acknowledges the assistance of Joana Mendes, Luís Seca, J. Peças Lopes, and Manuel Matos in the preparation of this report and work.

The authors are grateful to EDP Renewables North America LLC for providing wind farm data used in the analysis presented in this report.

The authors acknowledge the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, for funding the research presented in this report through its Wind and Water Power Program under contract DE-AC02-06CH11357.

ABBREVIATIONS

AIC	Akaike information criterion
ANN	artificial neural network
ARMA	autoregressive moving average
ARIMA	autoregressive integrated moving average
DFG	dynamic factor graph
DOE	U.S. Department of Energy
EM	expectation maximization
h	forecasting time step
HRRR	high-resolution rapid refresh
ITL	information theoretic learning
MAE	mean absolute error
MAP	maximum a posteriori
MSE	mean square error
MCC	maximum correntropy criterion
NBIAS	normalized bias
NMAE	normalized mean absolute error
NRSME	normalized root mean square error
NSDE	normalized standard deviation error
NOAA	National Oceanic and Atmospheric Association
NWP	numerical weather prediction
PDF	probability density function
RMSE	root mean square error
SCADA	supervisory control and data acquisition
U	east–west wind component
V	north–south wind component
VST	very short term
W2P	wind to power
WFA	Wind Farm A
WFB	Wind Farm B
WFC	Wind Farm C
WFD	Wind Farm D

WFIP	Wind Forecast Improvement Project
WPF	wind power forecasting
w.r.t.	with respect to

DYNAMIC FACTOR GRAPHS – A NEW WIND POWER FORECASTING APPROACH

ABSTRACT

Wind energy production is characterized by high uncertainty due to the random nature of wind speed, which is weather dependent. Thus, in power systems with a significant penetration of wind generators, the problem of studying the impact of wind farms on the performance of these systems has to be considered in detail and requires the availability of accurate wind power forecasting methods. In addition, the time horizons to be taken into account as references for such studies are becoming more and more variable (from a few hours to a few days) due to the needs of liberalized electricity markets.

This report presents a short outline of a new methodology for modeling and forecasting wind power generation on a very short-term time scale (up to 6 hours). The class of dynamic factor graph (DFG) models is described jointly with the algorithm for conducting fitting, inference, and forecasting within this framework. The DFG model class encompasses the well-known state-space models (sometimes referred to as Kalman filters) and allows a host of generalizations (e.g., higher Markov orders, as well as nonlinear state and observational functions).

Several experimental DFG models were applied to the problem of forecasting wind power generation for the very short term. The performance of these models was assessed against a benchmark persistence model and the best autoregressive model (autoregressive integrated moving average [ARIMA] model). In our preliminary experiments with DFGs, we observed that the lower Markov orders (namely, Markov order 1) achieved better results in terms of root mean square error (RMSE) and mean absolute error (MAE) improvements (relative to the persistence benchmark and to the ARIMA class) and fitting accuracy.

We present results for a forecast horizon of 0–6 hours for two actual wind farms in the Midwest United States (Wind Farm A [WFA] and WFB datasets) and for a wind farm in the South Central United States (WFC dataset). Data from WFA, WFB, and WFC were originally recorded at 10-minute time steps, but we also ran experiments at lower data frequencies. The tables that follow summarize the results for forecasting 0 to 6 hours ahead in terms of improvements over the persistence and ARIMA models for MAE and RMSE.

Average Improvement (0–6 hours) for DFG Model w.r.t Persistence Benchmark Model	MAE			RMSE		
Time Resolution	10 min	30 min	60 min	10 min	30 min	60 min
WFA	32.4%	32.3%	23.7%	28%	29.6%	20.7%
WFB	34.3%	34.5%	30.1%	31.3%	29.9%	29.7%
WFC	34.4%	29%	33.7%	30.6%	25.9%	30.7%

Average Improvement (0–6 hours) for DFG Model w.r.t. ARIMA Model	MAE			RMSE		
Time Resolution	10 min	30 min	60 min	10 min	30 min	60 min
WFA	3.8%	3.8%	7.6%	1.8%	1.6%	5.2%
WFB	2.5%	6.2%	15.9%	0.9%	2.5%	13.9%
WFC	3.5%	5.7%	12.6%	1.3%	3.1%	9%

According to both the RMSE and MAE measures, the improvement obtained by the first-order DFG models with respect to (w.r.t.) the ARIMA model class for the WFA, WFB, and WFC datasets recorded at a 60-minute frequency was always positive. This indicates that the DFG models provide superior forecasting performance when compared to the classical ARIMA approach. At a 30-minute data frequency, the average improvement obtained by the first-order DFG models w.r.t. the ARIMA model class for WFA, WFB, and WFC datasets was also always positive. However, some individual time horizons did lead to negative improvement values w.r.t. the ARIMA class. For the WFA dataset, those points were 30 minutes ahead (forecasting time step [h] = 1), 60 minutes ahead (h = 2), and 90 minutes ahead (h = 3) for the RMSE measure; they were 60 minutes (h = 2), 90 minutes (h = 3), and 120 minutes ahead (h = 4) for the MAE measure. For the WFB dataset recorded at a 30-minute frequency, only two time horizons led to negative improvement values w.r.t. the ARIMA class: h = 1 and h = 6 (30 and 180 minutes ahead) for the RMSE measure. Finally, for the WFC dataset, only one point led to negative improvement values w.r.t. persistence: h = 2 (60 minutes ahead) for the RMSE measure.

For the high-frequency datasets (10-minute time resolution), the advantage of the DFG approach over the classic ARIMA time series models was also consistent for all three wind farms. Strictly speaking, the average improvement w.r.t. the ARIMA class obtained by the DFG models was always positive for both measures (RMSE and MAE). However for some particular forecasting horizons, the ARIMA models outperformed the DFG ones. For the WFA dataset, those points were h = 1, 2, and 3 for the RMSE measure and h = 1 and 2 for the MAE measure. For the WFB dataset, those points were h = 2, 5, 6, 14, 30, and 31 (20, 50, 60, 140, 300, 310 minutes ahead) for the RMSE measure and h = 2 for the MAE measure. For the WFC dataset, no DFG *failure* point was found; that is, the DFG models outperformed the ARIMA class for all forecast horizons.

For the Midwest WFA dataset, we had access to external variables: numerical weather predictions (NWPs), which included U (east-west) and V (north-south) wind speed components measured at 10 and 80 m aboveground and pressure. Following previous studies, we selected the wind velocity components as explanatory variables to be introduced in the first-order DFG models, particularly the wind speed components measured 80 m aboveground. We combined the U and V wind speed components as a bivariate explanatory variable and measured the resulting improvement compared to the simplest DFG model (without external variables). For both RMSE and MAE measures, the improvement values increased for the first three forecast horizons. We

also introduced the wind speed modulus into the first-order DFG model; again we observed, for the first three forecast horizons, a significant increase in the forecast improvement.

Due to the lack of continuity in the NWP data and for comparison purposes, we used a wind to power (W2P) model in order to test the performance of the NWPs. Based on WFA experiments, we could affirm that even when the NWPs are not good enough to provide, by themselves, accurate wind power predictions for the very short term, they are able to increase the improvement of time-series-based models. We also tested the W2P model with NWP data for a second wind farm in the South Central United States (WFD), and we obtained better results than those for WFA.

In this study, we reached the following overall conclusions on the basis of our experiments with wind power forecasting for a very short-term time horizon:

- The DFG model was adapted in order to be applied to the very short-term wind power forecasting problem (0–6 hours). The DFG is within the general class of state-space models, and it provides more flexibility in representing time patterns (i.e., it enables the description of more complex time series patterns) than do the traditional time series models such as the ARIMA models. The results obtained for three real wind farms in the United States show that DFG models consistently outperform persistence and ARIMA models. Therefore, we recommend the adoption of state-space models like DFG for this type of problem.

- The inclusion of NWPs, combined with the past values of the time series, resulted in a slight improvement in the forecast accuracy (compared to that of a DFG model without NWPs) for the first three lead time steps. Nevertheless, we must emphasize that in our experiments, the main contribution to the improvement over persistence and ARIMA comes from using past values from the time series as input to the DFG model and not from the use of NWPs.

- The addition of NWP inputs, at the level of quality available for this study, to a W2P model, was not able to provide, by itself, enough improvement to the forecast accuracy over the level offered by a persistence model, for the first 5 hours for WFA and for the first 2 hours for WFD.

This page intentionally blank.

1 INTRODUCTION

With the current focus on the environmental sustainability of energy production, efficient integration of renewable energy sources into the electric power systems is becoming increasingly important. In Europe, several countries already have a high penetration of wind power (e.g., it ranges from 7% to more than 20% of electricity consumption in countries such as Germany, Spain, Portugal, and Denmark). The rapid growth in installed wind power capacity is expected to continue in the United States as well as in Europe [1].

However, as is well known, wind energy production is characterized by high uncertainty due to the random nature of the wind speed, which is weather dependent [2, 3]. Thus, in transmission and distribution systems with a significant penetration of wind generators, the problem of studying the impact of wind farms on the performance of these systems has to be considered in detail and requires the availability of accurate wind power forecasting methods. In addition, the time horizons to be taken into account as references for such studies are becoming more and more variable (from only a few hours to a few days) due to the needs of liberalized markets [4, 5, 6].

The variability inherent in wind makes wind power a fluctuating source of energy. Therefore, as the penetration of wind power in the electric power grid increases, the development of accurate forecasting tools to predict wind power is crucial for all electricity market participants, in order to reduce the economic and technical risks associated with the uncertainty in the production of wind power. In fact, wind power forecasting models enable better power system planning and operation (e.g., with regard to dispatch, scheduling, and unit commitment) and more competitive market trading [7, 8]. Because these activities have different time scales, their forecasting time period dictates the design of the prediction system. Accordingly, wind power forecasting (WPF) techniques are typically focused on three distinct time scales: very short term (up to 6 hours), short term (up to 2–3 days), and medium term (up to a week).

The relevant literature that has been published recently shows that many research projects have been conducted for the purpose of developing reliable wind-forecasting methods. These efforts have produced methods with different levels of accuracy for different applications. In general, the existing methods can be broadly classified as physical methods, statistical methods and approaches based on artificial neural networks (ANNs), and hybrid approaches [9].

The physical methods use the physical information involved (e.g., wind conditions at the height of turbine hubs, turbine power curves, weather conditions) to obtain an estimate of the wind power. These methods can also use meteorological information provided by complex numerical weather prediction models [1].

The statistical methods forecast either a wind-speed/power value ("point-forecast" methods) or a wind-speed/power probability density function ("pdf-forecast" methods). Both are obtained from statistical analyses of time series data. The ANN methods likewise try to determine the relationship between current wind power and past time series data [1].

Finally, hybrid approaches are combinations of different methods (e.g., physical and statistical approaches or alternative statistical methods). In general, physical methods are characterized by significant computational complexity and guarantee good performance mainly for short-term wind-power forecasting (i.e., several hours to a few days ahead). On the other hand, for very short-term (VST) wind-power forecasting (i.e., a few hours ahead only), statistical methods can provide accurate results while requiring significantly less computational effort [6]. The choice of a wind power forecasting method should be made by taking into account both the considered application and the related time horizon that is needed.

An extended review of the state of the art of wind power forecasting methods and, more specifically, of ways to deal with the VST time-scale problem can be found in References [1] and [10], respectively.

This report focuses on the VST time scale and proposes a novel WPF method based on a new class of statistical time-series models based on factor graphs. We test the new forecasting algorithms and evaluate the forecast accuracy compared with that of other methods at three wind farms in the United States. We also investigate how high-resolution weather forecasts, such as the ones from the U.S. National Oceanic and Atmospheric Administration (NOAA) High-Resolution Rapid Refresh (HRRR) model,[1] produced for the U.S. Department of Energy (DOE) Wind Forecast Improvement Project (WFIP),[2] may contribute to improved VST wind power forecasts.

This report is organized as follows. Section 2 gives a brief introduction to time series modeling. Section 3 describes the theory of dynamic factor graphs (DFGs). Section 4 describes the modeling assumptions of the DFG approach applied to the wind power forecast problem. Sections 5 through 8 present the results for real wind farms. Finally, Section 9 presents the concluding remarks and future work.

[1] http://ruc.noaa.gov/hrrr/.

[2] http://www.esrl.noaa.gov/psd/psd3/wfip/.

2 TIME SERIES

2.1 TIME SERIES MODELING WITHOUT HIDDEN VARIABLES

Time series modeling is motivated by a wealth of interesting problems; examples include forecasting (e.g., WPF), recovering missing data points (e.g., noisy signals in electrical engineering), learning a dynamic system (system identification), classifying subsequences, and estimating the likelihood of a sequence. This section provides a brief introduction to time series modeling. For a general discussion of the main concepts mentioned here, see Reference [11] or [12], for example.

One approach that is frequently used to model time series data is to assume that the current value Y_t of a given time series is a function *f* of the *p* previous values of *Y* up to some zero-mean random noise; that is,

$$Y_t = f\left(Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}\right) + \varepsilon_t$$

Eq. 1

The main limitation of this approach is that the "memory" of the model only goes as far as the number of *p* previous values used. This means that there is no memory of the full time series or of the long-term dependencies. This class of models may be satisfactory for handling time series that present simple stationary behavior, but it is usually not capable of capturing long-range dynamics. Models of this type are built under the so-called Markov assumption; that is, that the random variable Y_t conditional on the sequence $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}$ is independent from any random variable Y_τ with $\tau < t - p$. An issue that is immediately relevant is the choice of the order *p* for a particular time series. The traditional solution to this problem consists in choosing the order that minimizes some type of information criteria (e.g., the Bayesian information criterion or the Akaike information criterion).

2.2 TIME SERIES MODELING WITH HIDDEN VARIABLES

Traditional statistical time series models can be classified as one of two mutually exclusive classes: (1) models defined by observations or (2) parameter-driven models. Parameter-driven models are characterized by the presence of latent (or hidden) variables. Hidden variables *Z* are a way to summarize the history of a time series *Y*. A model with system memory or state information can be developed by the use of hidden variables. This memory is introduced as a dynamic relationship between the consecutive values $\ldots, z_{t-1}, z_t, z_{t+1}, \ldots$ of the hidden variables. The observed time series *Y* is considered to be generated by the hidden variables *Z*; that is, the observed variables are the result of the hidden process *Z*.

Two examples may clarify this discussion. The classical autoregressive moving average (ARMA) time series model

$$Y_t = c + \varepsilon_t + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Eq. 2

is an example of an observation-driven model. On the other hand, the class of state-space models adequately illustrates the general concept of parameter-driven models:

$$\begin{cases} Z_t = f(Z_{t-1}) + \varepsilon_t \\ Y_t = h(Z_t) + \omega_t \end{cases}$$

Eq. 3

The use of hidden variables is a natural way of handling the inherent incompleteness of time series data. There are three reasons why time series measurements are, in essence, incomplete. First, the observations are recorded at (discrete) sampling points, although the system being observed often evolves continuously over time. Second, almost invariably, all that is measured is merely a subset of the variables involved in the problem. Third, the process generating the observed time series may not be time-invariant (i.e., the mechanism generating the observations may be evolving through time). If this evolution is not so irregular as to render useless any attempt at forecasting, then it will almost certainly be necessary to keep track of the time trajectory of the system. In other words, a model with memory will be required. As mentioned, such a model can be obtained by including hidden variables.

3 DYNAMIC FACTOR GRAPHS

3.1 FORMALISM

A factor graph is a bipartite graph with two types of nodes, variables *Y* and *Z,* and factors *h* and *f* (which are functions of variables). The edges in a factor graph are such that each variable node can be directly connected only to factor nodes, and each factor node can be directly connected only to variable nodes.

Dynamic factor graphs or DFGs [13] are factor graphs specially tailored for handling sequences of random variables. The architecture of a DFG describing the structure of a general state-space model, like the one in Eq. 3, is presented in Figure 1.

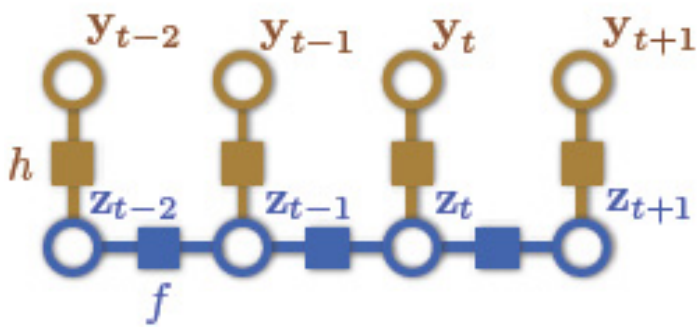


FIGURE 1 Architecture of a DFG that describes the structure of a general state-space model [13]

DFGs are more flexible than state-space models, however, because they allow for any conceivable dependence structure between observed and hidden variables. For example, consider the DFG architecture presented in Figure 2.

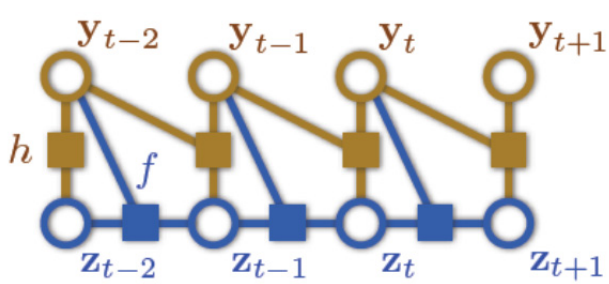


FIGURE 2 Architecture of a DFG with a more complex structure [13]

This architecture corresponds to the structure of the following class of time series models:

$$\begin{cases} Z_t = f(Z_{t-1}, Y_{t-1}) + \varepsilon_t \\ Y_t = h(Z_t, Y_{t-1}) + \omega_t \end{cases}$$

Eq. 4

where *f* and *h* are given functions and ε_t and ω_t are noise processes.

3.2 FACTORS USED

DFG models may also handle a situation in which one time series *Y* is considered the output of a system subject to an input sequence *X*. A particular architecture for such a case is presented in Figure 3.

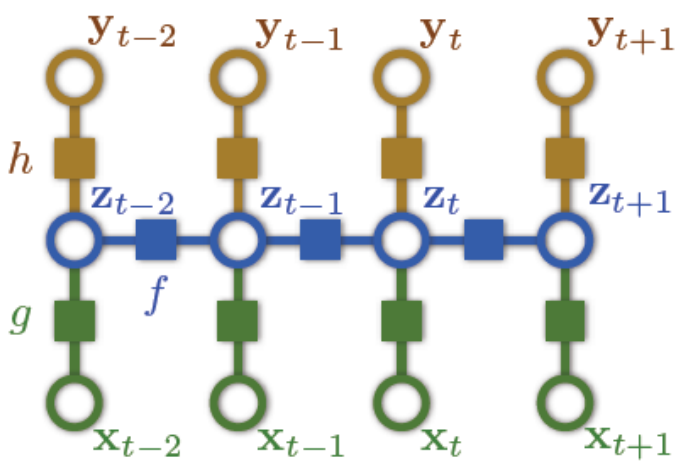


FIGURE 3 DFG structure for a case in which the time series *Y* is considered the output of a system [13]

This corresponds to time-series regression models and the study of the input-output relationship between two observed time series. This architecture describes the structure of the following class of time series models:

$$\begin{cases} Z_t = f(Z_{t-1}, X_t) + \varepsilon_t \\ Y_t = h(Z_t) + \omega_t \end{cases}$$

Eq. 5

The functions for all kind of factors (observational, dynamic, or exogenous) can be linear or nonlinear. A nonlinear function can help us obtain a more accurate description of the real world problem, but the nonlinearity increases the computational complexity and therefore the computational effort.

3.3 MAXIMUM LIKELIHOOD ENERGY-BASED INFERENCE

Once we have defined the architecture of the model involving hidden variables, inference on the historical sequence of hidden variables *Z* can be conducted. The aim is to find the *Z* sequence that optimally represents the observed variable *Y* (and *X*, if relevant) under the model.

3.3.1 Energy as Negative Log-Likelihood

Let us introduce the notion of *energy*, which is reviewed in References [14] and [15]. The notion of energy can be understood as resulting from the total prediction error, represented as $E(a_t, o_t)$ in Figure 4. Using the factor graph formalism in the logarithmic domain, the energy of the whole sequence of observed and hidden variables is a sum of energies of all the factors and is denoted as $E(Y, Z; X)$. Making the model parameters *W* explicit in the energy term, we write $E(Y, Z; X, \mathrm{W})$.

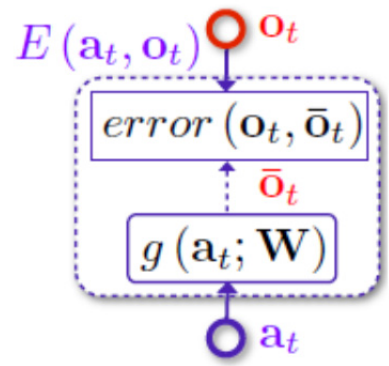


FIGURE 4 General description of a factor linking variables a_t and o_t through function g, with energy term $E(a_t, o_t)$ [13]

The total energy associated with the DFG is made proportional to the joint log-likelihood of observed and latent variables *Y* and *Z* given the inputs *X*:

$$E(Y, Z; X, W) \propto -\log P(Y, Z|X, W) + const$$

Eq. 6

Note that the energy presented does not by itself define a probability distribution because the normalization terms are unknown. To solve that problem, one would need to resort to the so-called Boltzmann distribution [14]:

$$P(Y, Z|X, W) = \frac{e^{-\beta E(Y,Z;X,W)}}{\int_{D_Y}\int_{D_Z} e^{-\beta E(Y',Z';X,W)} dY' dZ'}$$

Eq. 7

$$= \frac{e^{-\beta E(Y,Z;X,W)}}{\Gamma_{Y,Z}(X, W)}$$

Eq. 8

This distribution provides the maximum entropy distribution that is still compatible with the observations. The normalization constant $\Gamma_{Y,Z}(X, W)$ is called the partition function.

In order to evaluate the observed sequence Y, one has to marginalize $P(Y, Z|X, \mathrm{W})$ and Eq. 6 over all the values that the hidden sequence Z can take:

$$P(Y|X, W) = \int_{D_Z} P(Y, Z|X, W)\, dZ$$

Eq. 9

$$= \int_{D_Z} \frac{e^{-\beta E(Y,Z;X,W)}}{\Gamma_{Y,Z}(X, W)}\, dZ$$

Eq. 10

The evaluation of the integrals in Eq. 10 over all observed and hidden sequences is intractable for continuous variables under non-Gaussian distributions, or even when the distributions are Gaussian but the factors are nonlinear.

The use of the Maximum A Posteriori (MAP) approximation, proposed in Reference [2], which foregoes the full distribution in favor of its mode, permits one to derive:

i. $E(Y; X, W) = arg\, min_Z\, E(Y, Z; X, W)$ *Eq. 11*

ii. $arg\, min_Z\, E(Y, Z; X, W) = arg\, max_Z\, P(Z|X, Y, W)$ *Eq. 12*

The proof for this MAP approximation is given in Appendix A.

3.3.2 Inference of Latent Variables

For a given configuration of the parameters, the latent variables inference in a MAP setting is equivalent to finding the optimum of $E(Y, Z; X, W)$, which is the solution to the equation:

$$\frac{\partial E(Y, Z; X, W)}{\partial Z} = 0$$

Eq. 13

This can be achieved by using the well-known gradient descent algorithm. The gradients are back-propagated [16] from the energy modules in both directions, and each z_t is updated by summing up the contributions coming from all the factors to which it is connected. Until a convergence criterion is achieved, the gradient step is repeated.

3.4 EXPECTATION MAXIMIZATION-BASED LEARNING OF DFGS

3.4.1 Expectation Maximization Algorithm

Reference [17] first introduced the expectation maximization (EM) algorithm for the iterative computation of maximum-likelihood estimates when the observations can be viewed as incomplete data. The EM algorithm is directed at finding the parameters that maximize the objective function given an observed time series sequence. EM alternates between parameter estimation/learning (M-step) and latent variables inference (E-step), and can be understood as a gradient ascent algorithm applied to the likelihood. The main limitation of the EM algorithm is that it is only assured to converge to a local maximum (i.e., there is no guarantee that it will always achieve the global maximum).

The EM algorithm aims at maximizing the joint likelihood $P(Y, Z|X, W)$ of the complete data (i.e., observed plus hidden variables) specified by a model parameterized by the unknown vector W. Because the variables Z are latent, it tries to maximize the expectation of the log-likelihood with respect to the hidden data. The first step (E-step) consists in evaluating $E(\log P(Y, Z|X, W^{(k)}))$ given the current estimate $W^{(k)}$ of the parameters. The second step (M-step) consists in maximizing that quantity with respect to (w.r.t.) the parameters W; that is, assigning $W^{(k+1)} = \arg\max_{\mathrm{W}} E(\log P(Y, Z|X, W^{(k)}))$.

In Reference [18], an alternative justification of the EM algorithm is presented, in terms of free energy and entropy. In particular, the distribution $P(Y, Z|X, W)$, which is unknown, is replaced by an approximate distribution $Q(Y, Z|X, W)$ that is known, and during the E-step, instead of maximizing the expectation of $P(Y, Z|X, W)$, one maximizes the logarithm of $Q(Y, Z|X, W)$, which is proven to be a lower bound for the log-likelihood $\log P(Y, Z|X, W)$ [19].

3.4.2 Simplification and Approximation

Following Reference [13], we make use of the MAP approximation to the full distribution $P(Y, Z|X, W)$. Therefore, maximizing the conditional likelihood of the hidden variables is equivalent to minimizing the energy, as set out in Eq. 12.

3.4.3 Alternated E-Step and M-Step Procedure

The learning process in a DFG consists of adjusting the parameters W in order to minimize the sum of energies at each factor. It is possible to consider the introduction of the regularization terms $R_Z(Z)$ for constraining the Z sequence and $R(W)$ for regularizing the W factors, and, eventually, an $R_{CR}(Z)$ term for penalizing the dependency between each pair of Z components. This leads from the energy function to the loss function defined in Eq. 14.

$$\mathcal{L}(Y,Z;X,W) = \sum_{t=1}^{T}(\alpha E_d(t) + \beta E_o(t) + \gamma_Z R_Z(Z) + \gamma_W R(W) + \gamma_{CR} R_{CR}(Z))$$

Eq. 14

where $E_d(t)$ and $E_o(t)$ refer to the dynamic and observational energies defined in Eq. 15 and Eq. 16, respectively.

$$E_d(t) = E_d\left(W_d, z_{t-p}^{t-1}, z(t)\right) = \|\bar{z}(t) - z(t)\|_2^2$$

Eq. 15

$$E_o(t) = E_o(W_o, z(t), y(t)) = \|\bar{y}(t) - y(t)\|_2^2$$

Eq. 16

According to Eq. 12 and Eq. 15, the iterative procedure can be written as:

$$\mathcal{L}(Y,Z;\mathrm{X},\mathrm{W}) = \sum_{t=1}^{T}(\alpha E_d(t) + \beta E_o(t) + \gamma_Z R_Z(Z) + \gamma_W R(\mathrm{W}) + \gamma_{CR} R_{CR}(Z))$$

Eq. 17

$$\mathbf{E-STEP}: \tilde{Z} = arg\min_{Z} \mathcal{L}\left(Y,Z;X,\widetilde{\mathrm{W}}\right)$$

Eq. 18

$$\boldsymbol{M-STEP}: \widetilde{W} = arg\min_{W} \mathcal{L}\left(Y,\tilde{Z};X,W\right)$$

Eq. 19

Minimization of the loss function is done iteratively in an EM-like fashion, in which the state vector *Z* plays the role of auxiliary variables. The inference described in Eq. 18 can be considered as the E-step (state update) of a deterministic gradient-based version of the EM algorithm. During the parameter-adjusting M-step (weight update) described by Eq. 19, the latent variables are frozen. This means that we are back into the nonhidden variable framework and that we perform any kind of optimization to adjust *W*.

3.5 FEATURES

If the DFG model is given by the following:

$$\begin{cases} Z_t = f\left(Z_{t-1}, \cdots, Z_{t-p}; X_t; W_D\right) + \varepsilon_t \\ Y_t = g(Z_t; W_o) + \omega_t \end{cases}$$

Eq. 20

where ε_t and ω_t are the innovation processes (or noise), then the structure of the loss function is as shown in the following.

The dynamic energy at time *t* is given by

$$E_d(t) = \left\| Z_t - f(Z_{t-1}, \cdots, Z_{t-p}; X_t; W_D) \right\|_2^2$$

Eq. 21

The contemporary observational energy is

$$E_o(t) = \| Y_t - g(Z_t; W_o) \|_2^2$$

Eq. 22

The regularization term on *Z* can take several different forms, but it can be specified, in particular, as a quadratic variation penalty to enforce smoothness on the latent sequence:

$$R_Z(Z) = \sum_{t=1}^{T} (Z_t - Z_{t-1})^2$$

Eq. 23

The parameter regularization term can be specified as $R(W) = \|W_O\|_1 + \|W_x\|_1 + \sum_{c=1}^{\#components} \left\| W_{d_C} \right\|_1$.

If the *Z* sequence variable has more than one hidden component, then a correlation term or dependency penalty term could be introduced; it is given as follows:

$$R_{CR}(Z) = \sum_{t=1}^{T} \sum_{i \neq j}^{\#components} \left| corr\left(z_i(t), z_j(t)\right) \right|$$

Eq. 24

where $corr(.)$ corresponds to the linear correlation coefficient between each pair of components of the hidden variable *Z*.

3.6 PERFORMANCE MEASURES

After we have trained the DFG (by using the training set), the trained DFG is used to generate forecasts over a distinct test (or validation) set. The forecasting performance can be assessed in both absolute and relative terms by using several different measures. The performance measures used here for WPF are (1) the evaluation of the prediction error using both the mean absolute error (MAE) and root mean square error (RMSE), (2) the computation of the standard deviation of the prediction error, and (3) the generation of nonparametric confidence intervals at a level of at least 75%.

This page intentionally blank.

4 APPLICATION TO TIME SERIES MODELS AND DYNAMIC SYSTEMS

This section outlines the numerical experiments that were used to perform the training and to test the DFG models. The empirical results of these experiments are reported in Section 5.

As depicted in Eq. 4 and 5, a DFG model is expressed by two equations as in a state-space model. In the general case, there is:

$$\begin{cases} y_t = h(Z_t; W_o) + \omega_t \\ z_t = f\left(Z_{t-1}, \cdots, Z_{t-p}; X_t, W_D, W_X\right) + \varepsilon_t \end{cases}$$

Eq. 25

where Z_t refers to the hidden/dynamic variable at time *t*; X_t is the explanatory/exogenous vector at time t; and W_o, W_D, and W_X are the model parameters corresponding to the observational, dynamic, and regression components. ω_t and ε_t are the noise processes corresponding to the observational and state equations.

At present, we consider both the observational and dynamic functions to be linear. Most of our dataset is made up of wind power measurements recorded at time steps of 10, 30, and 60 minutes.

Combining the different elements mentioned yields the following system:

$$\begin{cases} y_t = W_o Z_t + \omega_t \\ Z_t = W_{d1} Z_{t-1} + W_{d2} Z_{t-2} + \cdots + W_{dp} Z_{t-p} + W_x X_t + \varepsilon_t \end{cases}$$

Eq. 26

The loss function in the more general case is as follows:

$$\mathcal{L}(Y, Z; X, W) = \sum_{c=1}^{\#components} \left(\alpha \sum_{t=1}^{T} E_d(t) + \beta \sum_{t=1}^{T} E_o(t) + \gamma_z R_Z(Z) + \gamma_W R(W) \right)$$

Eq. 27

The algorithm for training the DFG model and conducting inference in this context uses the gradient descent algorithm. It is therefore necessary to obtain the gradient of the loss function.

$$\nabla \mathcal{L}(Y, Z; X, W) = \left(\frac{\partial \mathcal{L}}{\partial Y}, \frac{\partial \mathcal{L}}{\partial Z}; \frac{\partial \mathcal{L}}{\partial W_D}, \frac{\partial \mathcal{L}}{\partial W_O}, \frac{\partial \mathcal{L}}{\partial W_X} \right)$$

Eq. 28

Note that the gradient of the loss function does not have any component expressing variation with respect to the exogenous variable *X*. This is because all inference is conducted conditionally on the observed values of *X*. The derivation of the gradient and initializations of the latent variable can be found in Appendix B.

The gradient descent algorithm is used for updating the estimates of both the model parameters W and the latent variables Z:

$$\begin{cases} Z^{(k+1)} = Z^{(k)} - \eta_Z \dfrac{\partial \mathcal{L}}{\partial Z} \\ W^{(k+1)} = W^{(k)} - \eta_W \dfrac{\partial \mathcal{L}}{\partial W} \end{cases}$$

Eq.29

A successful update depends largely on the value used for the learning coefficient represented as η_Z and η_W in Eq. 29. Typical values for the learning coefficient lie in the range [0,0.3], depending on the total range of the data in use. We implement a reduction in the learning rates η_Z and η_W in order to guarantee a fast convergence and also to avoid cases of divergence.

TABLE 1 Parameters and their values

Parameter	Value
α, β	0.5
γ_W	1
γ_{CR}	0 or 10
γ_Z	1
Epochs	10
η_{WD}	0.001
η_{WX}	0.01
η_Z	0.1
$\sigma_{dynamic\ innovation}$	1

There are several degrees of freedom (i.e., parameters) in this algorithm that have an impact on its convergence. The main parameters are presented in Table 1 jointly with the selected values. We arrived at these values by trial and error over many experiments; we started our trials on the basis of the suggestions made in Reference [13].

After training a particular DFG model with a specific Markov order, we tested its forecasting ability by using several relevant metrics: RMSE improvement compared with other methodologies (e.g., persistence and ARIMA class models), error bias, standard deviation, and a nonparametric confidence interval for the prediction error. In presenting the results, we concentrate mostly on the improvement achieved w.r.t. both the persistence and ARIMA models:

$$Improvement = 100 \times \frac{RMSE(persistence\ or\ ARIMA) - RMSE(DFG)}{RMSE(persistence\ or\ ARIMA)}$$

Eq. 30

where RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(forecast(i) - observed(i))^2}$$

Eq. 31

The improvement can also be computed with respect to the MAE measure, where MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|forecast(i) - observed(i)|$$

Eq. 32

4.1 MODEL TRAINING APPROACH

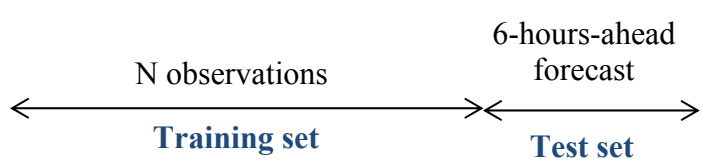
The model training approach employed makes use of a large initial training dataset for estimating the hidden sequence *Z*. Then the forecast for the next 6 hours (the test set) is made. The new training set is then considered to be the test set used in the previous estimation, and the necessary initial conditions are taken to be the last previously estimated values of the *Z* sequence. This represents "online" learning with a sliding-window approach. The forecast for the next 6-hour test set is again performed. This process is repeated until the entire dataset is exhausted.

In our experiments, we use an initial training set that covers between 1 month and 3.5 months of observations (depending on the total amount of data in the available datasets).

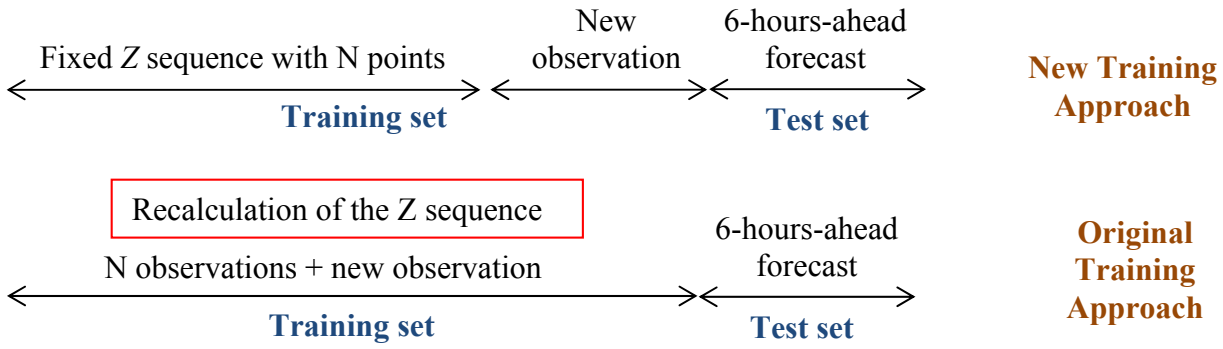
From our initial experiments with DFG models, we were able to observe that the best RMSE and MAE improvements (w.r.t. the persistence benchmark and to the ARIMA models) were obtained by models with low Markov orders. Accordingly, the experiences presented here use DFG models with Markov order $p = 1$.

Two training approaches (original and new) were developed and tested in this project. Both approaches have the following main steps:

Step 1: The training dataset with N data points of observations is chosen (1 month to 3.5 months). A forecast for the following 6 hours (test set) is made. The estimated *Z* sequence is fixed.



Step 2: The training set is now the fixed sequence of Z from the former training set plus the new observation (i.e., N + 1 observations). The initial conditions for the Z process over the training set are obtained from the estimated (fixed) *Z* sequence, and the Z value for the N + 1 observation is calculated. The forecast for the next 6 hours is made.



Note that for the original training algorithm, instead of simply estimating one new value for the *Z* sequence, it is necessary to estimate N + 1 values (re-estimating the complete Z sequence over the previous training set and estimating the new values over the new training set). This is very time consuming, since the size of the training set is continuously increasing. Therefore, in this report, we propose a new training approach that is computationally more efficient.

Step 3: Repeat step 2 until the entire test dataset is used.

The next three sections (5, 6, and 7) present the performance of the DFGs against both the well-known ARIMA class and the persistence benchmark over three datasets: Midwest Wind Farms A and B (WFA and WFB) and the South Central Wind Farm C (WFC).

5 EMPIRICAL RESULTS: MIDWEST WIND FARM A (WFA)

The wind power data used in this chapter are from a large wind farm located in flat terrain in the Midwest United States. The wind farm was divided in two "sub-wind farms" named Wind Farm A (WFA) and Wind Farm B (WFB). Data from both were originally recorded in 10-minute time steps. In this section, only the data from WFA are used. Results for WFB are presented in Section 6.

In order to empirically test this new modeling methodology, we fit the first-order linear DFG described in the previous section to three time series of WFA wind power generation recorded at sampling frequencies of 10, 30, and 60 minutes, respectively. The DFG model was updated for every time step (i.e., 10, 30, and 60 minutes) as new observations of wind power generation were made available by the supervisory control and data acquisition (SCADA) system.

The complete dataset (i.e., generated power as measured by the wind farm's SCADA) made available for this project corresponds to the period between January 2, 2009, and February 20, 2010.

In order to assess the behavior of the multi-step-ahead prediction error generated by the DFG, we ran the training algorithm as described in Subsection 4.1. From the 10-minute-frequency dataset, we took 5,000 data points (1 month of observations) as an initial training set. From the 30-minute-frequency (re-sampled from the 10-minute time series by calculating the average) dataset, we selected an initial training set with 5,000 observations (approximately 3.5 months of observations). Finally, from the dataset recorded at a sampling frequency of 1 hour (re-sampled from the 10-minute time series by calculating the average), we used 2,000 data points (i.e., around 3 months of observations) as an initial training set. The multi-step-ahead, out-of-sample forecasting horizon covers 6 hours (i.e., 36, 12, and 6 data points for the 10-, 30-, and 60-minute-frequency datasets, respectively). Before fitting the DFG model, we applied differences to the original wind power measurements first, and then we normalized the resulting data. The normalization was done by subtracting the sample mean and dividing the resulting differences by their standard deviation. The ARIMA models were fitted to the same transformed data. Unless otherwise stated in the text, the remaining period of the dataset was used to test and compare the forecasting models.

5.1 10-MINUTE-FREQUENCY DATASET

The original time series of wind power production contains 59,688 data points. From these, we selected an initial training set of 5,000 data points. Using this wind power dataset, we estimated a first-order DFG model and subsequently generated a multi-step, out-of-sample forecast over a horizon of 36 data points (6 hours). We compared the resulting forecasting errors with what can be achieved by using the so-called persistence model (i.e., using the last observation as the prediction for the next point — the optimal forecasting strategy for a Martingale process) as well as the best ARIMA class (obtained using the "auto.arima" function of the "forecast" package for the R language [20]).

Figure 5 shows the average RMSE improvement w.r.t. the persistence benchmark that was attained by the estimated DFG model. The DFG model with Markov order 1 presents a mean improvement of 28.03% over the 6-hour forecasting window.

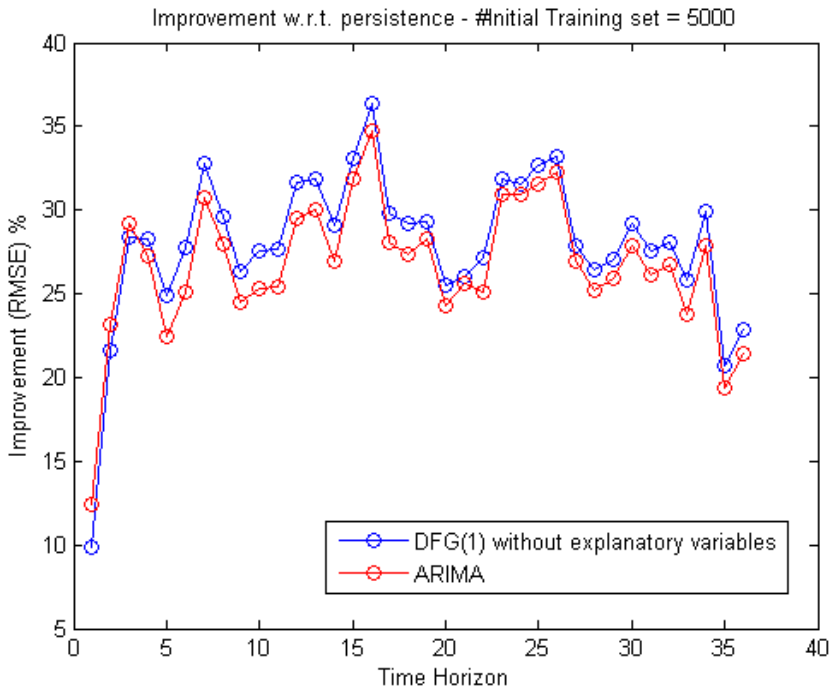


FIGURE 5 WFA for 10-minute- time step: RMSE improvements w.r.t persistence for the DFG models with Markov order 1 and the ARIMA model, with an initial training set with 1 month of observations (5,000 data points)

Figure 6 presents the MAE improvement w.r.t. the persistence benchmark that was attained by the estimated DFG model. The DFG model with Markov order 1 presents a mean improvement across the 6-hour forecasting window of 32.43%.

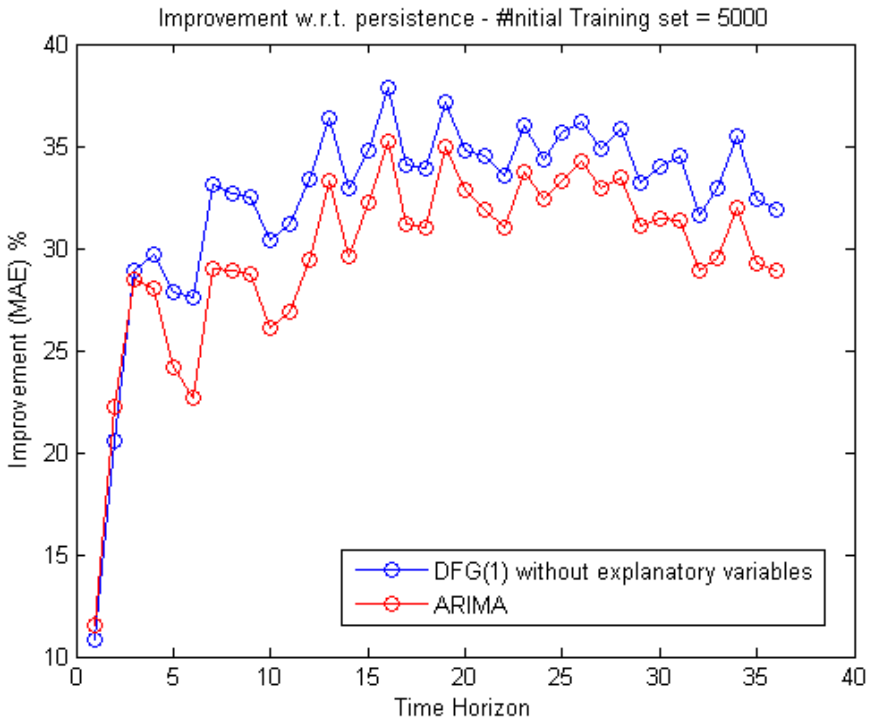


FIGURE 6 WFA for 10-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and the ARIMA model, with an initial training set with 1 month of observations (5,000 data points)

Besides comparing the forecasting performance of the DFG models with the persistence benchmark, we decided to use a less trivial class of time series models to help further assess the behavior of this new forecasting approach. Accordingly, we employed the class of autoregressive integrated moving average (ARIMA) models. By using the same training approach and after applying the same data transformation as used for the DFG models, we selected the ARIMA model that minimizes the Akaike information criterion (AIC). This particular ARIMA model was then used to forecast the next 36 observations over the test set. We conducted a search among all ARIMA models, starting from a pure white noise ARIMA(0,0,0) model up to an ARIMA(12,0,12) model. The "best" model was an ARIMA (3,0,1) model.

In addition to indirectly comparing the forecasting improvement of the DFG and ARIMA classes w.r.t. the persistence model, we show the direct improvement achieved by the DFG w.r.t. the ARIMA class in Figures 7 and 8. Figure 7 focuses on the RMSE, and Figure 8 focuses on the MAE. As could be expected after seeing the results from Figures 5 and 6, the RMSE improvement of the DFG models w.r.t. the best ARIMA model is positive from the fourth forecasting step until the thirty-sixth forecasting step. Similarly, the MAE improvement is positive from the third forecasting step onward. The maximum RMSE improvement was around 3.6% and was achieved for the forecasting step h = 6 (i.e., 1 hour ahead). With respect to the MAE improvement, its maximum (of around 6.4%) was also achieved at h = 6.

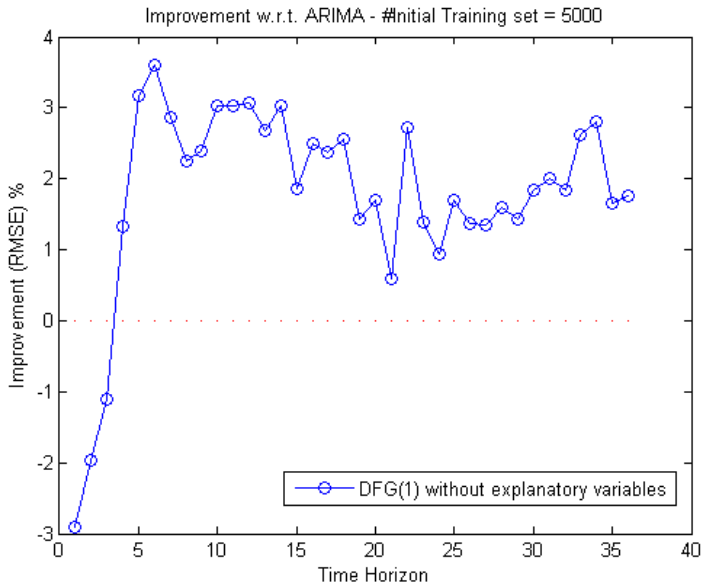


FIGURE 7 WFA for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

FIGURE 8 WFA for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

5.2 30-MINUTE-FREQUENCY DATASET

This dataset contains 19,895 data points and was obtained by aggregating the original dataset recorded at a 10-minute frequency. As we did for the first dataset, we initialized the State Sequence Z with a large training set: namely, 2,000 data points (i.e., 1 month of observations) for a DFG with Markov order $p = 1$. Again, we compared the forecasting performance of the DFG models with both the persistence benchmark and the minimum AIC ARIMA model.

For this sampling frequency, the advantage of the DFG models w.r.t. the best ARIMA model is less significant; the first 3 or 4 (out of 12) forecasting steps present a negative improvement. Figure 9 shows that the DFG model provides a larger RMSE improvement w.r.t. persistence than does the ARIMA model for the forecasting steps between 4 and 12. The average MAE improvement w.r.t. persistence obtained by the DFGs was larger than its ARIMA counterpart for all the forecasting steps except for the second and third (i.e., 1 hour and 1.5 hours ahead), as shown in Figure 10.

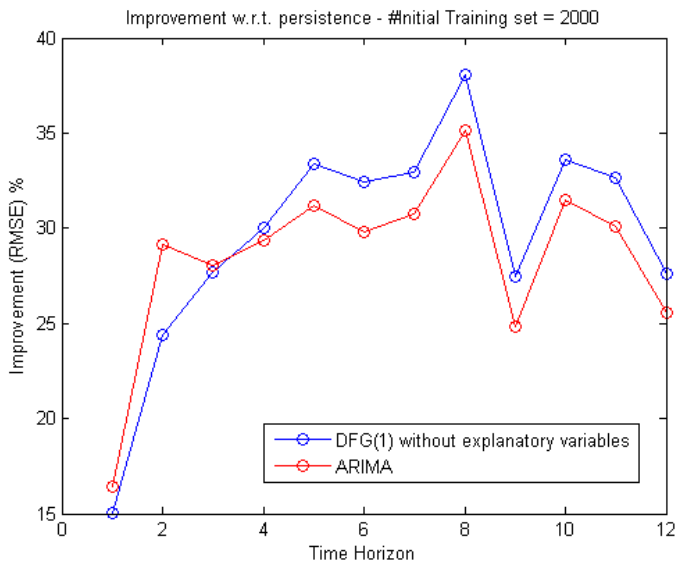


FIGURE 9 WFA for 30-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

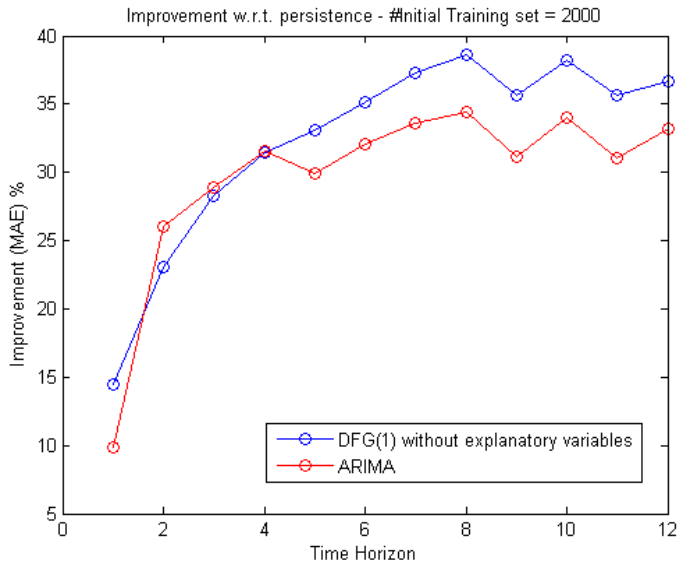


FIGURE 10 WFA for 30-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

Figures 11 and 12 show the improvement obtained by the DFG models w.r.t. the best ARIMA model. For most of the forecasting steps, that improvement is positive, reaching a maximum of 4.5% for h = 8 for the RMSE measure and 6.6 % for h = 11 for the MAE measure.

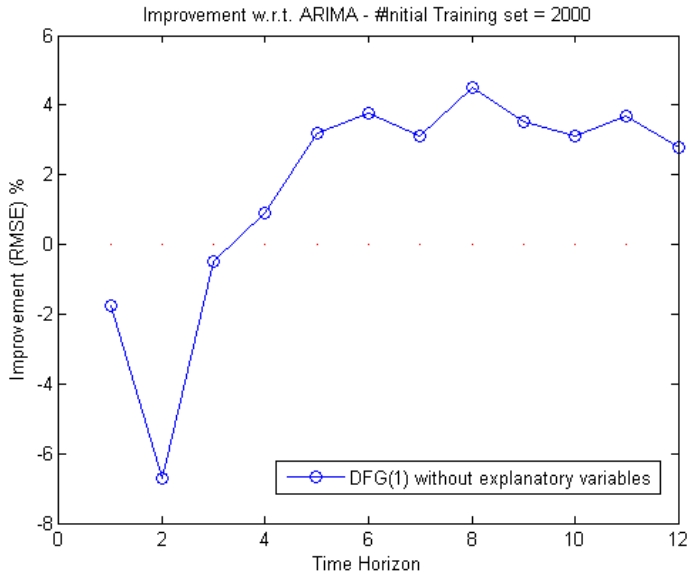


FIGURE 11 WFA for 30-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

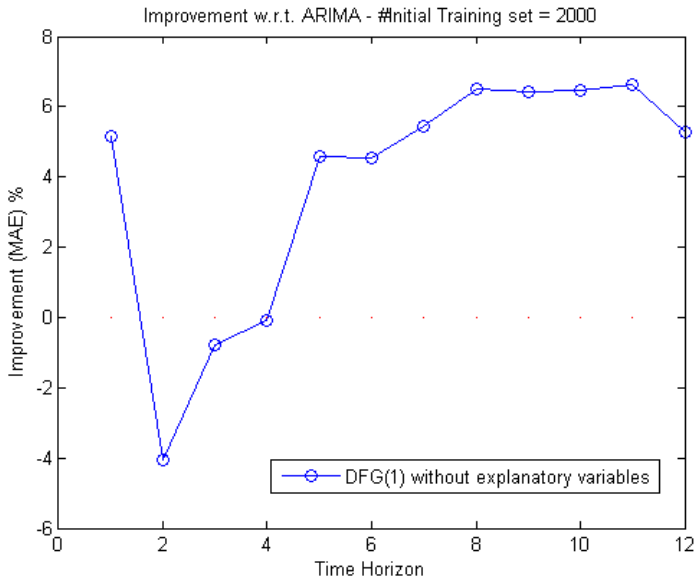


FIGURE 12 WFA for 30-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

5.3 1-HOUR-FREQUENCY DATASET

The last wind power dataset contains 9,947 data points and was similarly obtained by time aggregation. The size of the initial training set was 2,000 data points, which corresponds to approximately 2.8 months of observations. For this sampling frequency, the results show a clear-cut advantage of the DFG approach over the ARIMA class of time series models. The average RMSE improvement w.r.t. persistence obtained by the DFG model, 20.71%, surpassed what could be achieved by using the AIC-minimizing ARIMA (16.35%). Likewise, w.r.t. the average MAE improvement over persistence, the DFG model (with 23.73%) surpassed the best ARIMA model (with 17.48%). The comparison between the RMSE improvements (w.r.t. persistence) obtained by the DFG and ARIMA models is shown in Figure 13, while Figure 14 shows the corresponding comparison between MAE improvements.

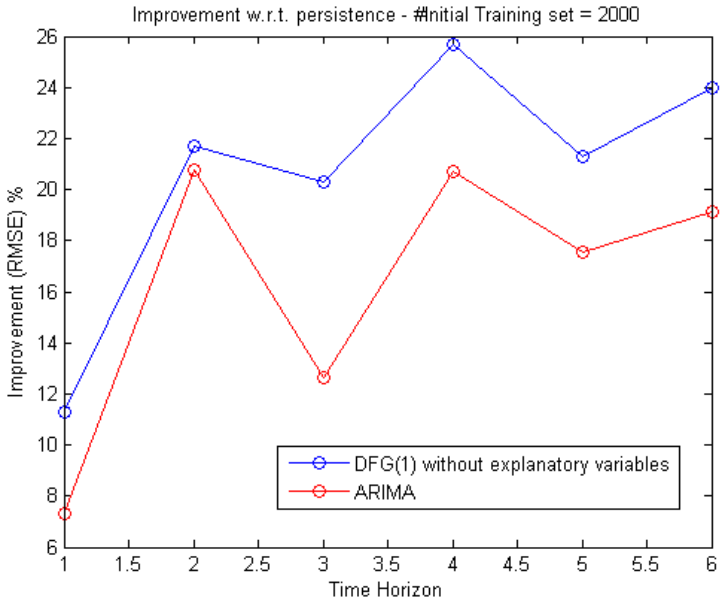


FIGURE 13 WFA for 60-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

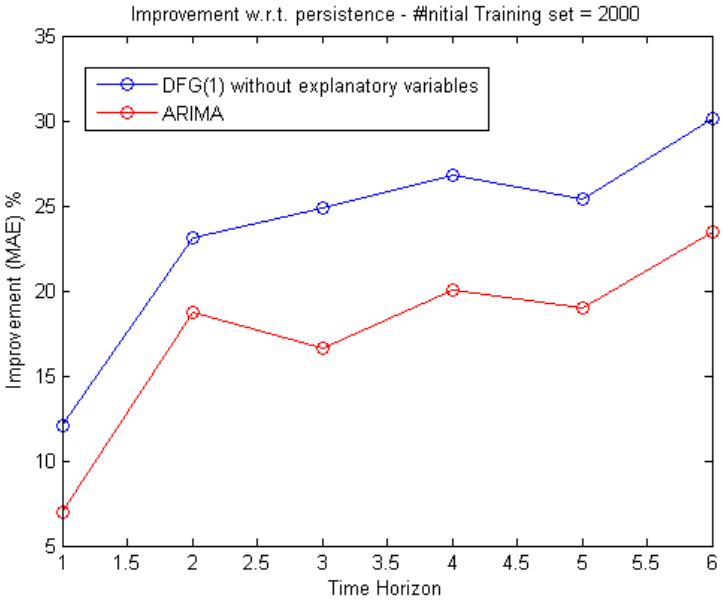


FIGURE 14 WFA for 60-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

Figure 15 presents the direct RMSE improvement achieved by the DFG model over the best ARIMA model, and Figure 16 shows a similar picture with regard to the MAE improvement measure. The improvement of the DFG models w.r.t. the best ARIMA model is positive for all the forecasting horizons for both the RMSE and MAE measures.

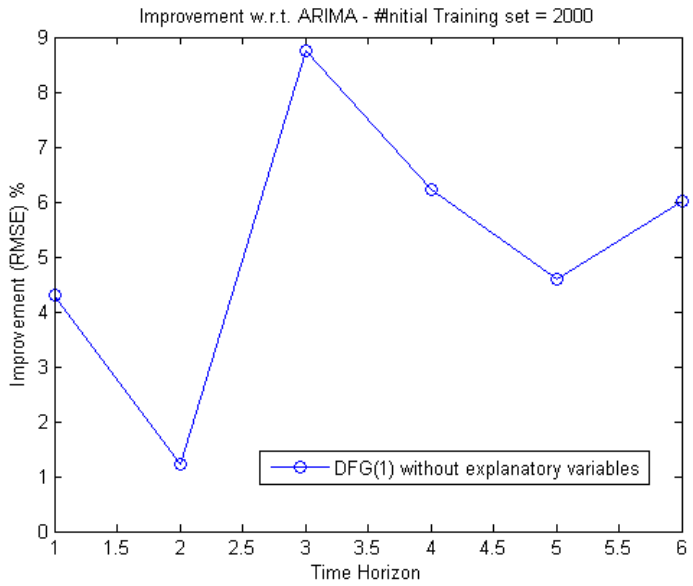


FIGURE 15 WFA for 60-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

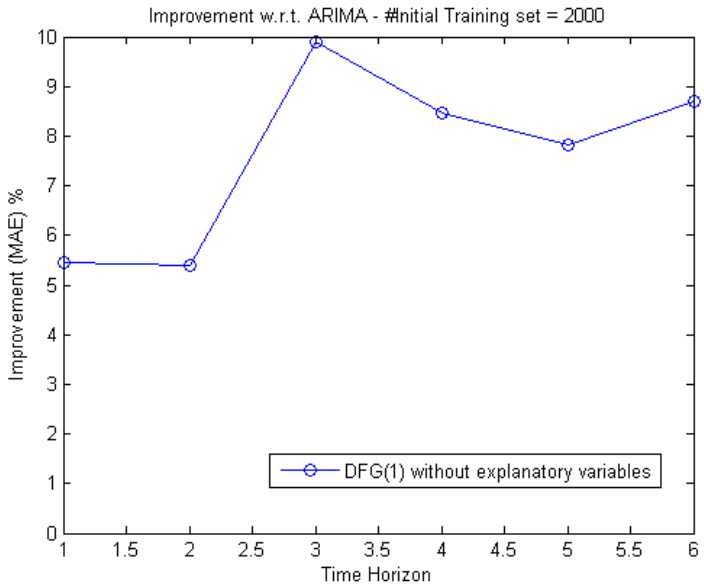


FIGURE 16 WFA for 60-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

5.4 MIDWEST WIND FARM A FOR ANOTHER TIME PERIOD

This subsection presents results for the same wind farm (WFA) but at a more recent time period for which we have access to numeric weather predictions (NWPs) from the HRRR model. The wind farm dataset now corresponds to the period between September 1, 2010, and August 31, 2012. Hence, the wind power dataset contains 17,079 data points measured at a 60-minute frequency. We have NWPs for only 6,372 hourly wind power measurements, however. In order to test the explanatory variables included in the DFG model, we based our results on the subset with 6,372 data points from December 1, 2011, to August 31, 2012. For this subset, we used an initial training set of 4,000 data points, which covers approximately 5.6 months of observations. The trained DFG model was then used to obtain forecasts up to 6 hours ahead. The DFG model still had a first Markov order (p = 1).

Here the results are presented first without considering the explanatory NWP variable. Figures 17 and 18 show the improvement of the DFG and ARIMA models over the persistence benchmark, which is positive for both methods. However, the DFG model obtained larger mean improvements: 25.48% for the DFG and 22.92% for the ARIMA under the RMSE measure (Figure 17), and 30.7% for the DFG and 24.12% for the ARIMA under the MAE measure (Figure 18).

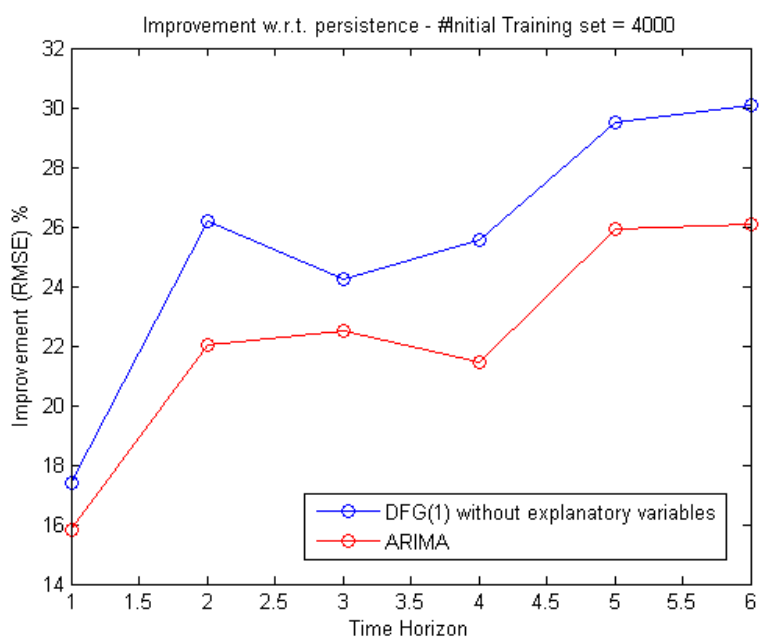


FIGURE 17 WFA for 1-hour time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

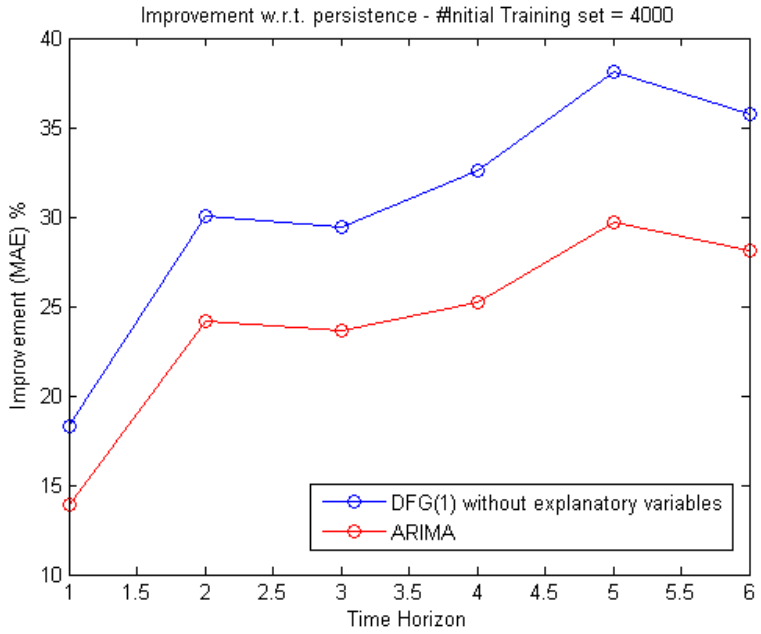


FIGURE 18 WFA for 1-hour time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

Figures 19 and 20 present the improvement results for the DFG over the ARIMA model. These are positive for all forecasting steps, with a maximum of 5.4% for h = 6 for the RMSE measure and 11.9% for h = 5 for the MAE measure. The DFG model is clearly more accurate than the persistence benchmark and ARIMA models.

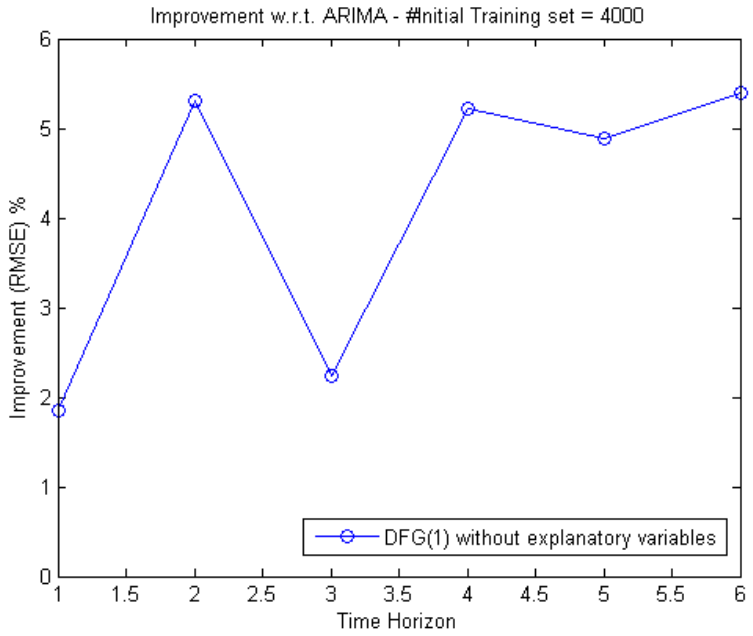


FIGURE 19 WFA for 1-hour time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

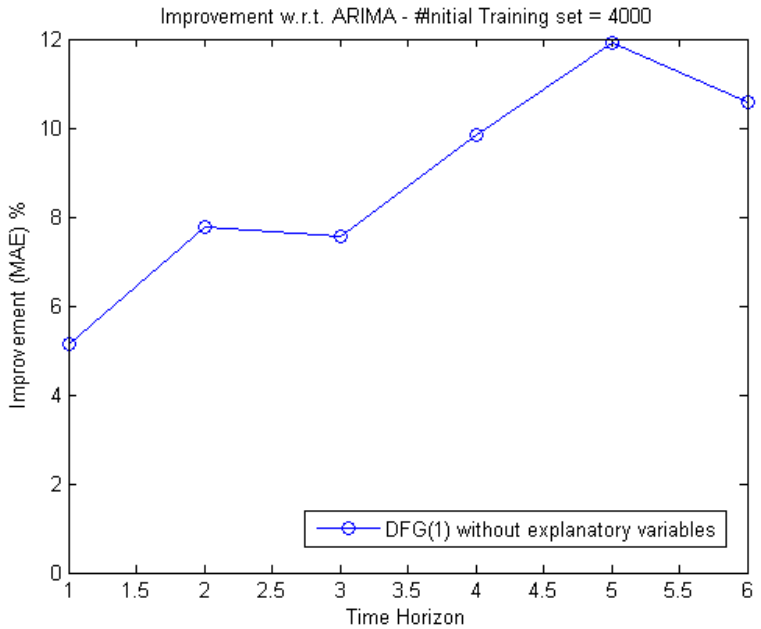


FIGURE 20 WFA for 1-hour time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

5.4.1 Including Explanatory Variables

To study the effect of additional explanatory variables in the DFG model, we introduced two different explanatory variables. First, we considered the U (east-west) and V (north-south) components of the wind from the NWPs produced by the HRRR model for DOE's WFIP. Second, we considered the forecasted wind speed modulus (i.e., calculated from the U and V components).

When the explanatory variables were first introduced into the DFG models, there were numerical instability and convergence problems. We found that the number of data measurements of the initial training set could not be made as large as the ideal desirable number because the computational problems were not manageable. Therefore we used 12 days and 12 hours (or 300 hourly data points) as the initial training set. Since the initial training set was not large enough, the update of the W factors constraint (see Subsection 4.1) was relaxed.

Figures 21 and 22 compare the improvement w.r.t. persistence for the first-order DFG model with no explanatory variables with that for the first-order DFG model with the NWP wind speed components and wind speed modulus introduced separately into the model. The figures show that both explanatory variables lead to an increase in the DFG improvement up to time horizon h = 4 for both RMSE and MAE measures. Hence, the use of these explanatory variables increases the accuracy of wind power forecasts up to 4 hours ahead.

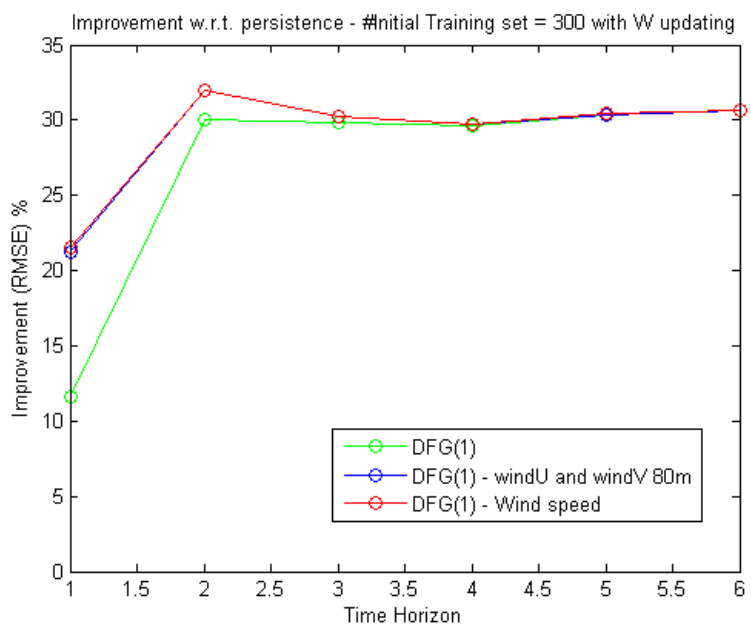


FIGURE 21 WFA for 1-hour time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 using the two components of the wind speed and the wind speed modulus as explanatory variables

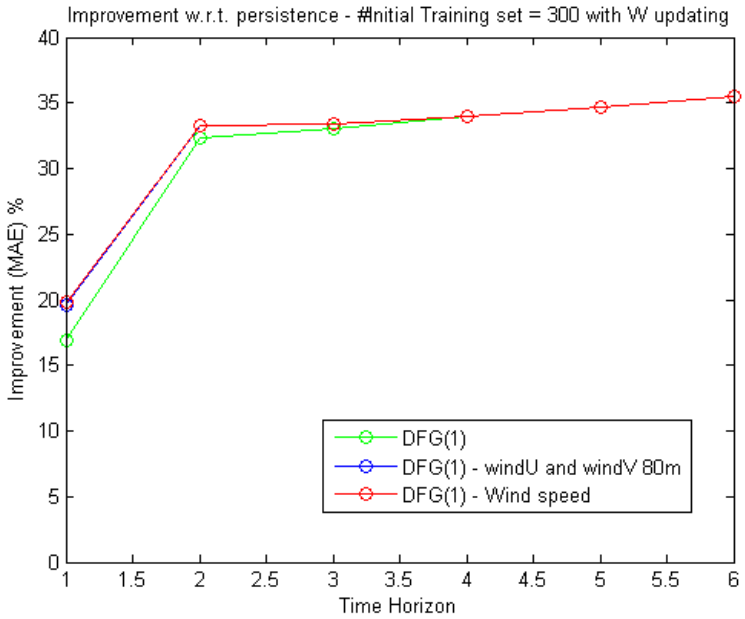


FIGURE 22 WFA for 1-hour time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 using the two components of the wind speed and the wind speed modulus as explanatory variables

We varied the number of data points used in the initial training set. However, the best framework that obtained the maximum improvement was for 300 initial data points with Z factors being updated, as explained in Subsection 4.1 (results are shown in Figures 21 and 22).

Finally, it is important to stress that the DFG model with NWPs also uses the past values of the time series; thus, this model can be seen as an autoregressive model with exogenous input. These results show that NWPs combined with past values can provide some improvements in forecast accuracy. Again, this empirical finding corroborates the previous result, pointing to the short memory and volatile nature of the wind speed. However, these results should be validated with data from other wind farms.

6 EMPIRICAL RESULTS: MIDWEST WIND FARM B (WFB)

The wind power dataset used in this section is also from the large-scale wind farm located in flat terrain in the Midwest United States. This section concentrates on the data from the second part of the wind farm analyzed in the previous section, labeled Wind Farm B (WFB). The dataset (i.e., generated power as measured by the wind farm's SCADA) corresponds to the period between January 2, 2009, and February 20, 2010. A striking feature of the time series that is analyzed here is that it involves long periods of time when the wind power output was not measured. Missing information is a common feature of wind power data and could be caused by several factors, including equipment failures, communication problems, and outages. To overcome this problem, the DFG model was applied only to periods without missing data; the blank periods were ignored.

6.1 10-MINUTE-FREQUENCY DATASET

The WFB dataset contains 21,999 data points. We used an initial training set of 5,000 data points, which corresponds to 1 month of observations. The trained DFG model was used to forecast for 36 data points (i.e., 6 hours ahead). The following results are from a DFG with Markov order 1.

The results for WFB are shown in Figures 23 through 26. The DFG models with Markov order $p = 1$ attained a larger average RMSE improvement than what was achieved by the best ARIMA model: 31.26% versus 30.62%, respectively. With regard to MAE improvement, these values were 34.26% and 32.54%, respectively.

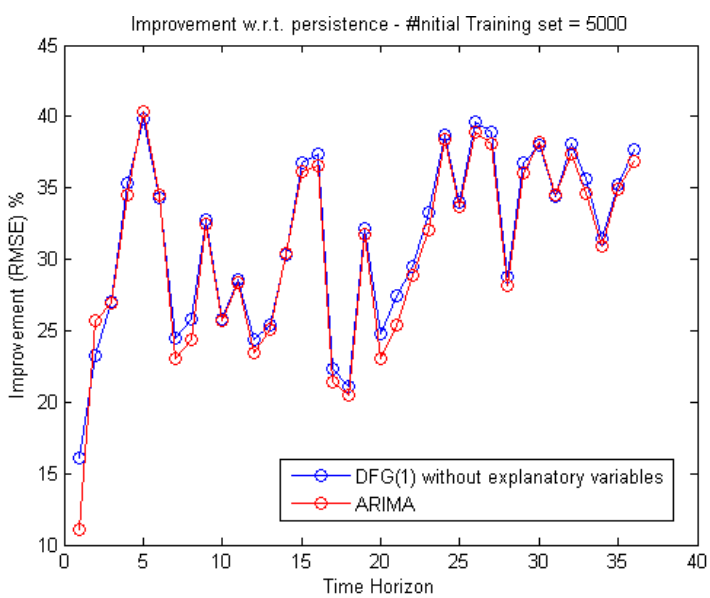


FIGURE 23 WFB for 10-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

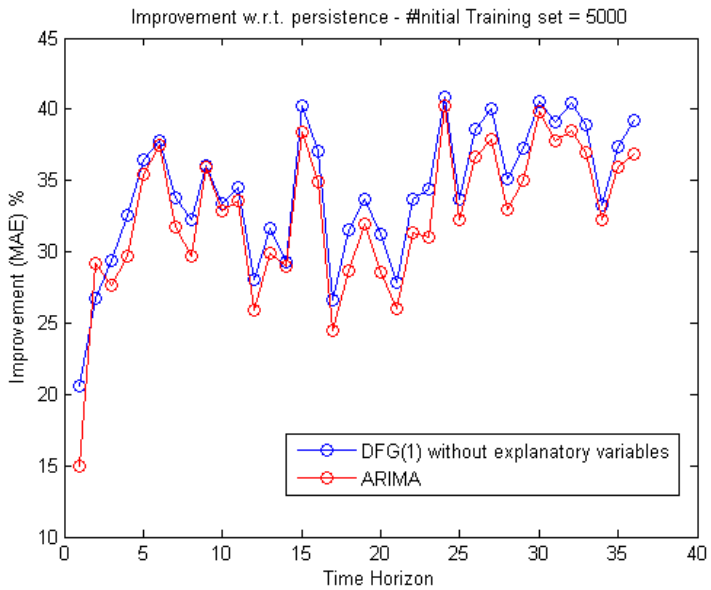


FIGURE 24 WFB for 10-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

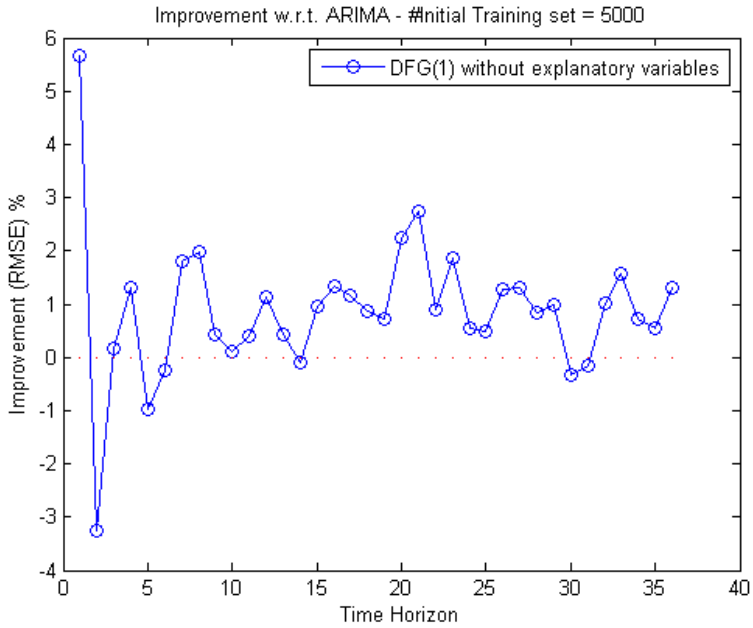


FIGURE 25 WFB for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

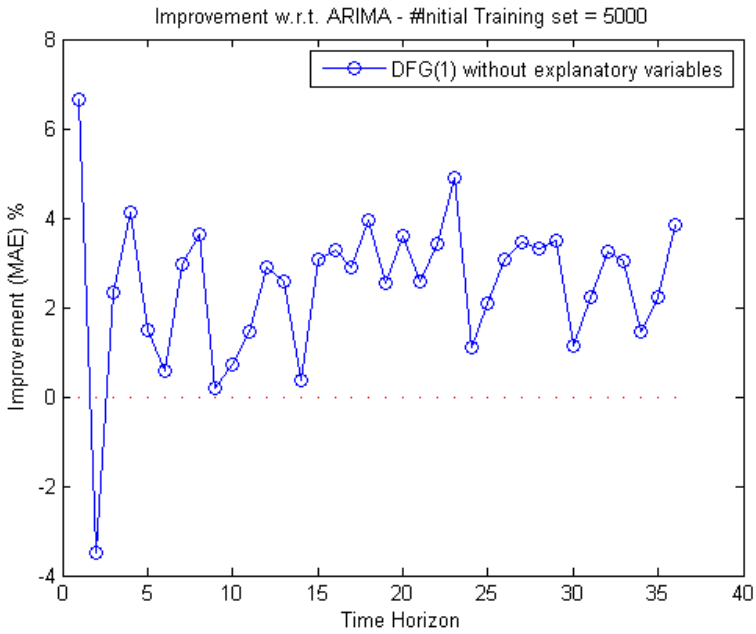


FIGURE 26 WFB for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

In Figures 25 and 26, it is possible to see the DFG improvement w.r.t. ARIMA model for each forecasting step. With regard to the RMSE measure (Figure 25), for the forecasting steps h = 2, 5, 6, 30, and 31, the DFG model leads to a negative improvement w.r.t. the ARIMA model. With regard to the MAE measure (Figure 26), the forecasting step h = 2 was the single case for which the DFG model could not outperform the ARIMA model. So, in general terms, these results show that for this dataset, the DFG approach outperforms the best ARIMA model, although the improvements are less significant than the improvements for WFA are for this frequency.

6.2 30-MINUTE-FREQUENCY DATASET

This dataset contains 7,332 data points and was obtained by aggregating the original dataset. For this case, we used an initial training set with 5,000 data points (equivalent to approximately 3.5 months of data). Similar to what was seen for the same dataset recorded at the 10-minute frequency, it is possible to see that a significant advantage is associated with using the DFG approach over the best ARIMA model. Figures 27 and 28 show that the improvement over the persistence benchmark attained by the best ARIMA model is significantly smaller than the improvement attained by the DFG approach. This is true for both the RMSE and MAE measures. Nevertheless, some forecasting horizons had lower RMSE improvement rates: namely the forecasting horizons h = 1 and h = 6.

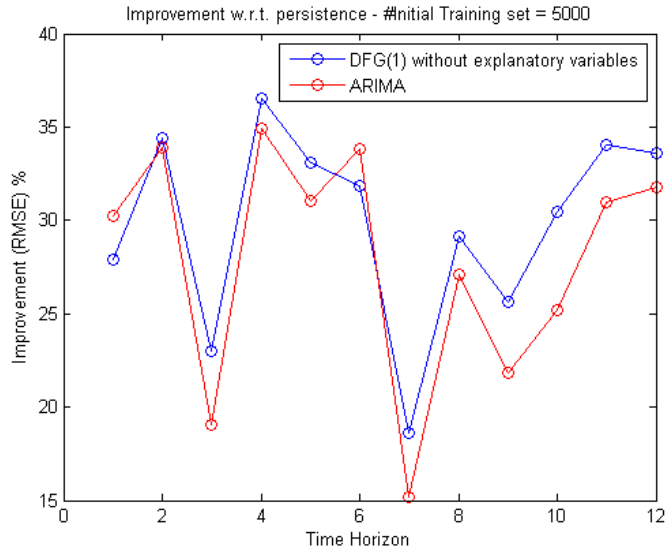


FIGURE 27 WFB for 30-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

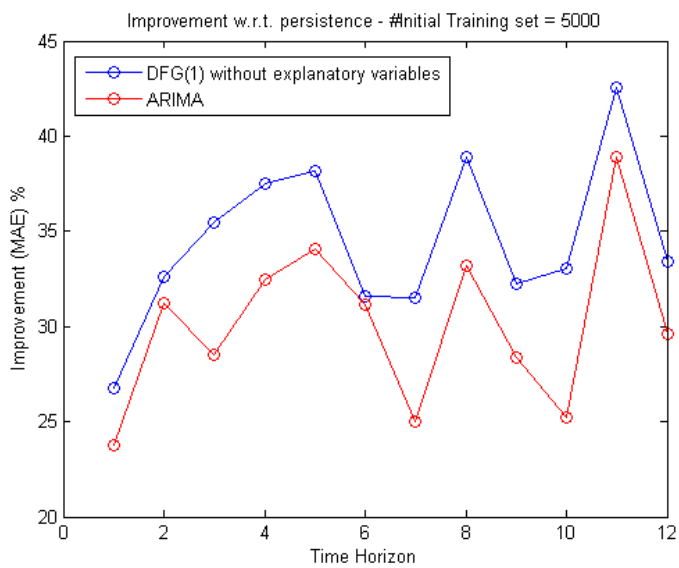


FIGURE 28 WFB for 30-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

Accordingly, when both the RMSE and MAE improvements for the DFG model w.r.t. the best ARIMA model are computed, Figures 29 and 30 show that the corresponding values are positive in general, except for the forecasting steps h = 1 and h = 6 for the RMSE measure. Under the MAE measure, there is an improvement for all forecasting time steps.

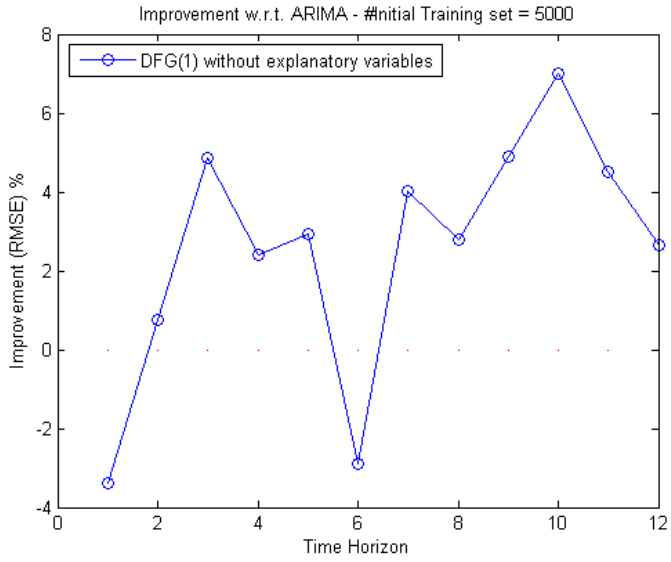


FIGURE 29 WFB for 30-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

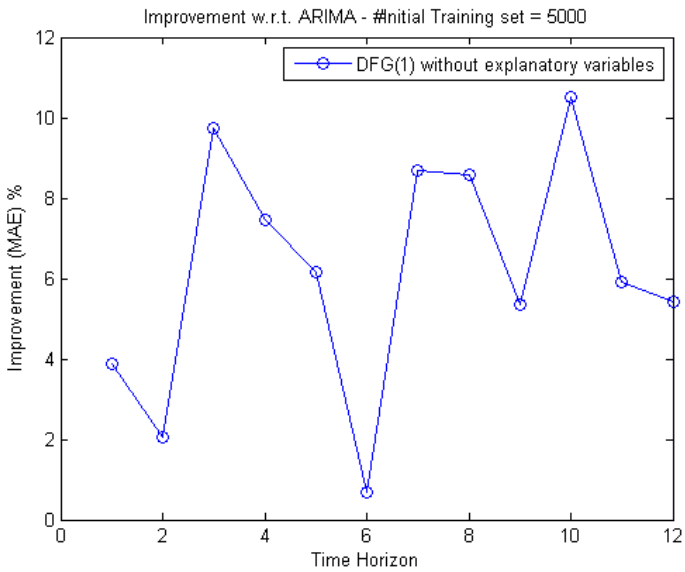


FIGURE 30 WFB for 30-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

6.3 1-HOUR-FREQUENCY DATASET

The last wind power dataset of the WFB dataset contains 3,665 data points and was obtained in a similar fashion, by the time aggregation of the dataset recorded at a 10-minute frequency. For this time series, we used an initial training set that covers approximately 2.8 months, or 2,000 data points.

The DFG model yielded a larger average RMSE improvement over the persistence benchmark than did the best ARIMA model (29.71% for the first-order DFG, and 18.21% for the best ARIMA; see Figure 31), and it likewise yielded a larger average MAE improvement (30.11% for the first-order DFG, and 16.84% for the best ARIMA; see Figure 32). The advantage of the DFG model over the best ARIMA is clear across the 6-hour forecast horizons; it can be seen in Figures 33 and 34. Similar to what happened for the WFA dataset recorded at a 1-hour frequency, the DFG approach clearly outperformed the class of ARIMA models; the relative improvement is even more pronounced for WFB. The DFG forecasts are, for the most part, smoother than the observed time series and ARIMA forecasts. Hourly data show less variability than do their higher-frequency counterparts, making their results more predictable. In general, the observed advantage of DFG models over ARIMA models confirms that the former are more flexible than the latter.

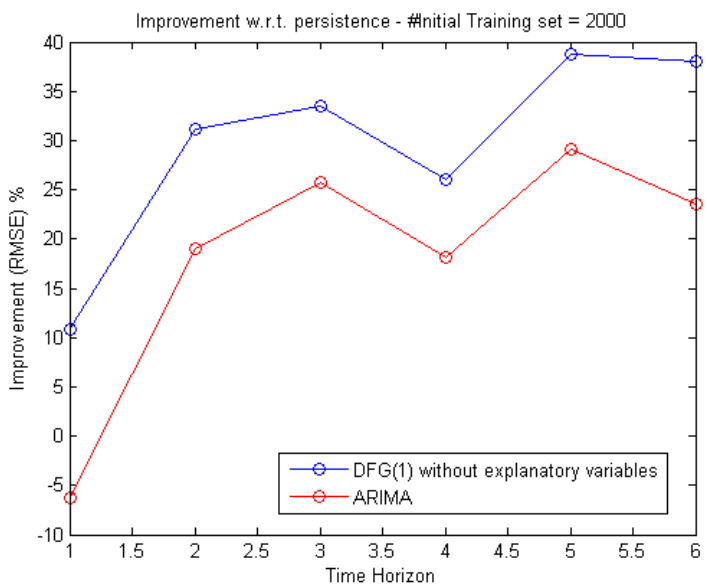


FIGURE 31 WFB for 60-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1

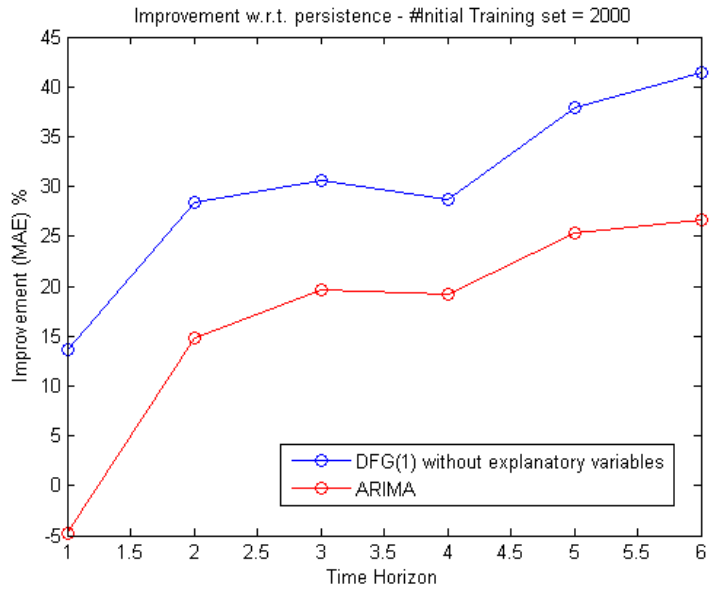


FIGURE 32 WFB for 60-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1

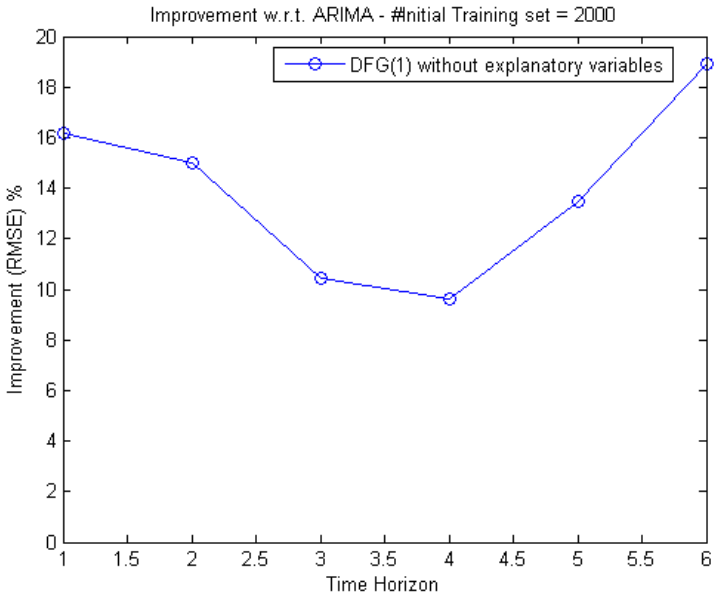


FIGURE 33 WFB for 60-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

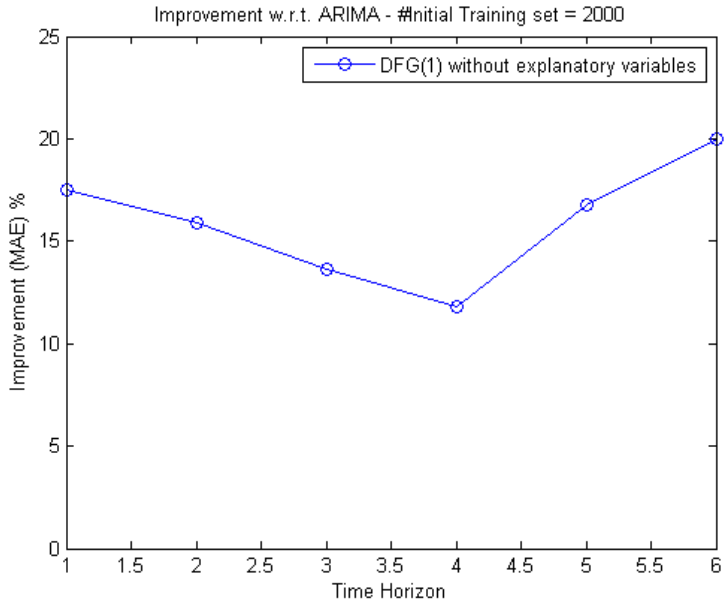


FIGURE 34 WFB for 60-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

Finally, note that we did not perform experiments with NWP explanatory variables for WFB because there were frequent gaps and missing data in the HRRR weather predictions to which we had access from the WFIP. The missing NWP data made it difficult to conduct experiments with statistical forecasting algorithms, since such models require relatively long periods of consistent data for both training and testing.

7 EMPIRICAL RESULTS: SOUTH CENTRAL WIND FARM C (WFC)

The wind power data used in this section are from a wind farm located in the South Central United States. The complete dataset available for WFC corresponds to the period between September 1, 2009, and August 31, 2012. In order to empirically test the new DFG forecasting methodology, we fit the first-order linear DFG to three time series of WFC wind power generation recorded at sampling frequencies of 10, 30, and 60 minutes, respectively. The results are presented in the following subsections.

7.1 10-MINUTE-FREQUENCY DATASET

The time series sequence of the wind power data for WFC contains 105,263 data points. For this dataset, we used an initial training set of 5,000 data points, which covers 1 month of observations. The trained DFG model was used to forecast the 36 data points ahead (i.e., the next 6 hours). The following results were obtained from a DFG model with Markov order $p = 1$.

The DFG approach achieved very good results for 6-hour-ahead forecasting for WFC: For all the forecasting steps, the DFG model had a higher improvement than did the ARIMA class. Figures 35 and 36 show that the improvement over the persistence benchmark attained by the best ARIMA model is surpassed by the DFG approach. Accordingly, Figures 37 and 38 show that the improvement of the DFG over the ARIMA class is positive for all forecasting steps. This result is in contrast to the results for the WFA and WFB datasets recorded at the 10-minute frequency, for which the best ARIMA model outperformed DFG for some forecasting time steps. Hence, DFG performed very well for WFC at the 10-minute frequency.

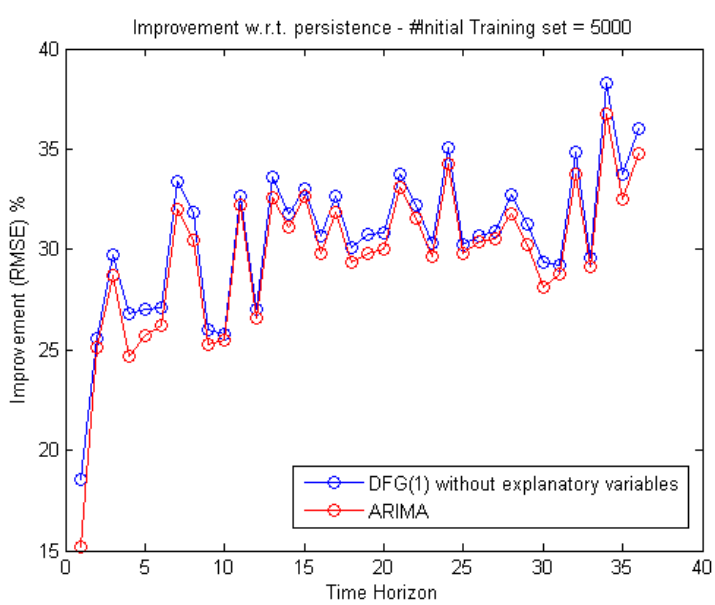


FIGURE 35 WFC for 10-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

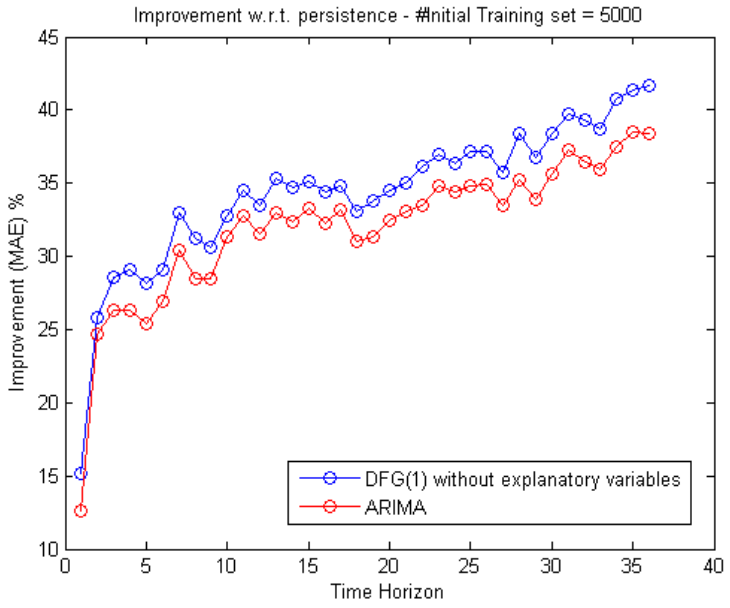


FIGURE 36 WFC for 10-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

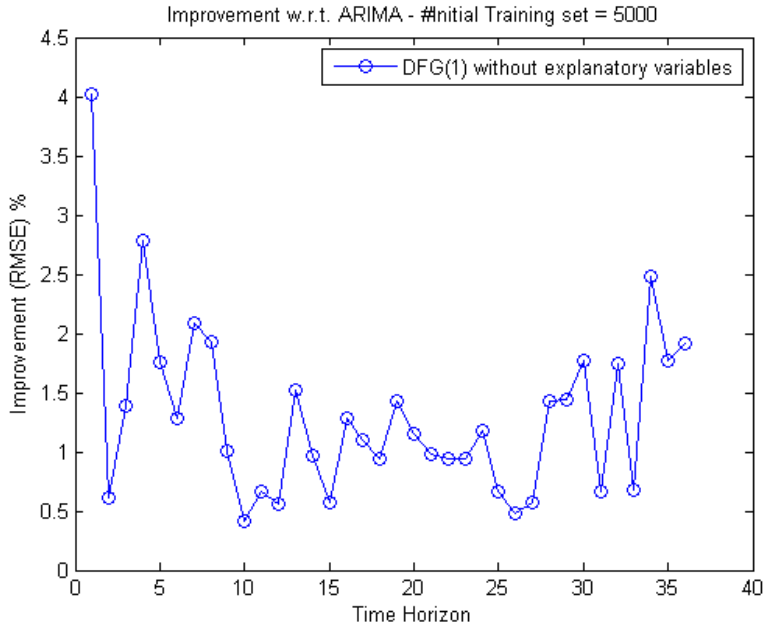


FIGURE 37 WFC for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

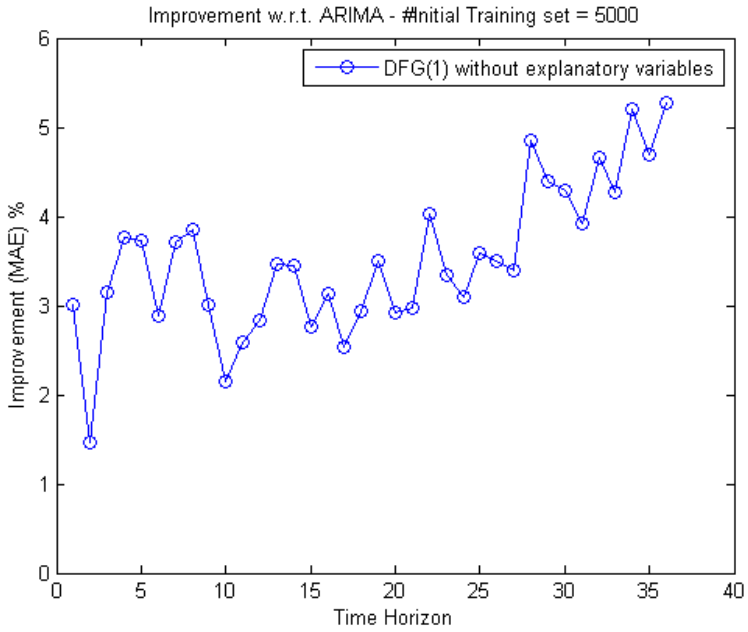


FIGURE 38 WFC for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

7.2 30-MINUTE-FREQUENCY DATASET

This dataset contains 35,087 data points. Here we used an initial training set of 2,000 data points, or approximately 1.4 months of observations. Our objective is to forecast the next 6 hours of production data (i.e., 12 data points into the future). The results that follow are from a DFG model with Markov order $p = 1$.

The results for this dataset are quite similar to what was obtained with the time series recorded at a 10-minute frequency. There is only one forecasting step (h = 2) for which the RMSE improvement achieved by the DFG model was lower than what the ARIMA model could achieve. The average RMSE improvement over the persistence benchmark was 25.87% for the DFG model and 23.43% for the ARIMA model (Figure 39). The MAE improvement was 29.04% for the DFG and 24.77% for the ARIMA model (Figure 40).

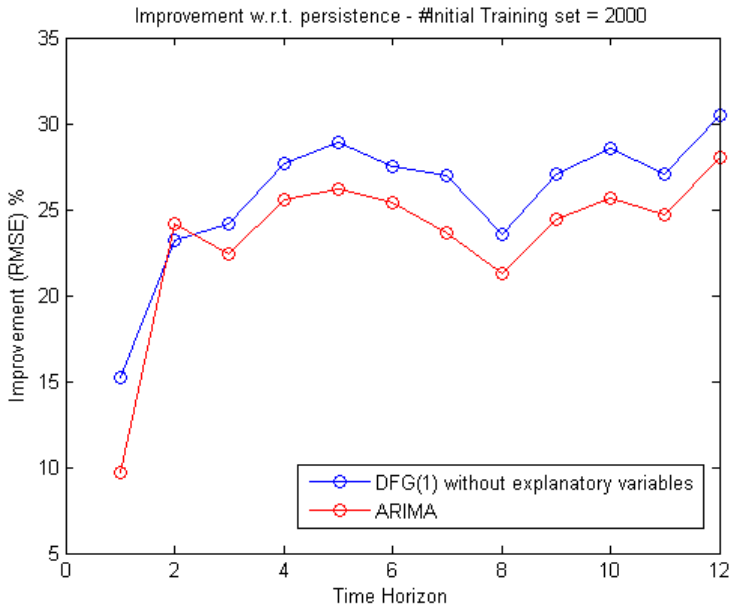


FIGURE 39 WFC for 30-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

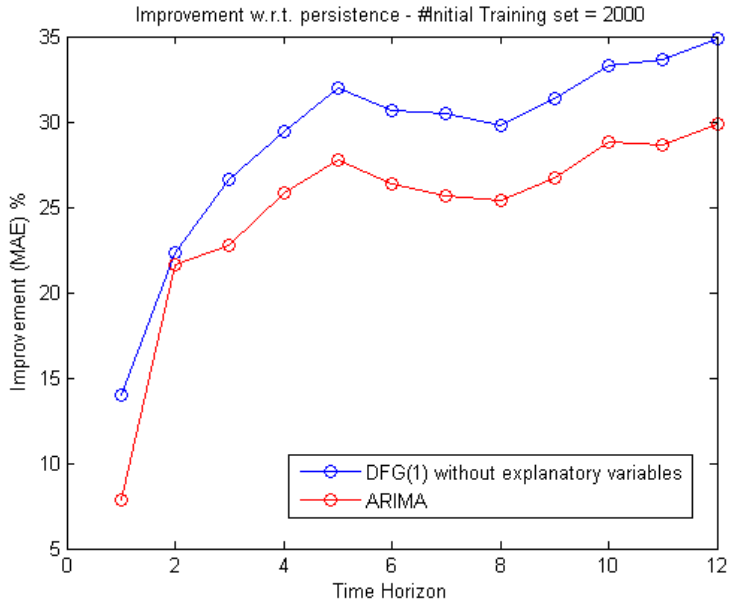


FIGURE 40 WFC for 30-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

Figures 41 and 42 show the improvement of the DFG model w.r.t. the ARIMA class. In agreement with the improvement w.r.t. persistence attained by the DFG and ARIMA models, Figures 41 and 42 show positive improvements; the only exception is the forecasting step h = 2 in the RMSE case.

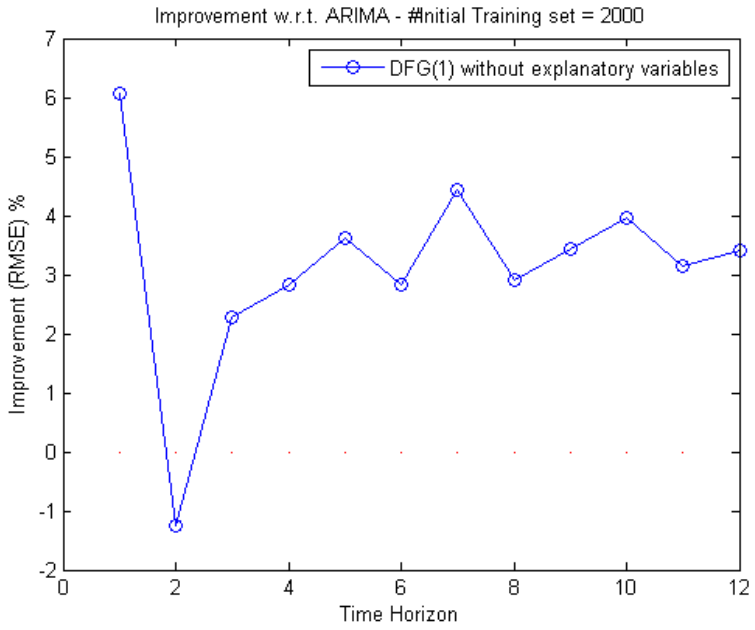


FIGURE 41 WFC for 30-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

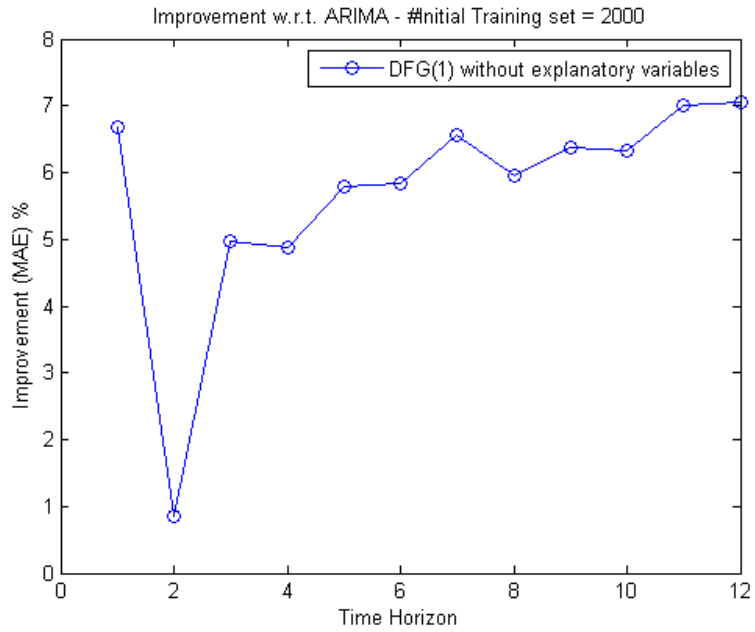


FIGURE 42 WFC for 30-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

7.3 1-HOUR-FREQUENCY DATASET

This dataset contains 17,543 data points. Here we used an initial training set of 2,000 data points, which covers approximately 2.8 months of observations. The trained DFG model was used to forecast the next 6 data points (i.e., a 6-hour window). The following results are from a DFG model with Markov order $p = 1$.

Again, the DFG approach proved to be the best forecasting technique, as it previously was for the WFA and WFB datasets recorded at this frequency. Figures 43 and 44 present the improvement w.r.t. the persistence benchmark. The average improvement was 30.70% for the DFG model and 23.99% for the ARIMA model for the RMSE measure; for the MAE measure, the average improvement was 33.7% for the DFG model and 24.44% for the ARIMA model. Figures 45 and 46 show that the improvement over the ARIMA model is always positive.

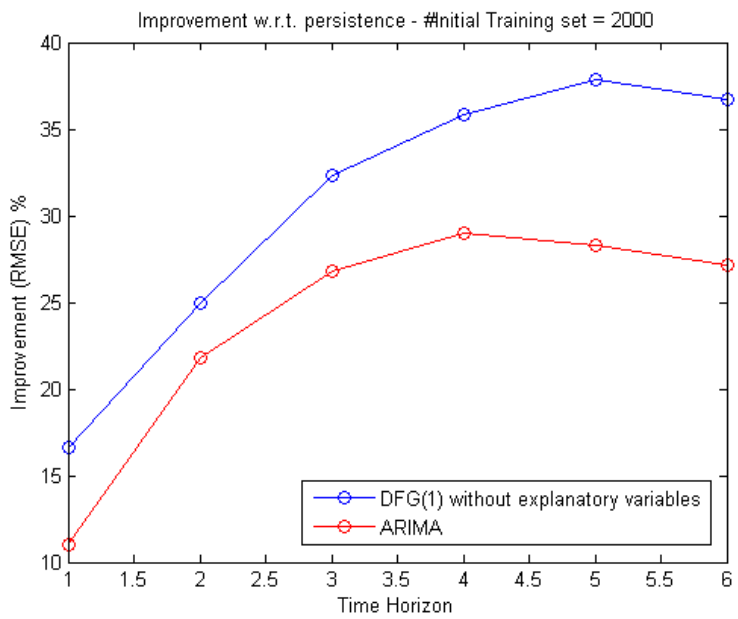


FIGURE 43 WFC for 60-minute time step: RMSE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

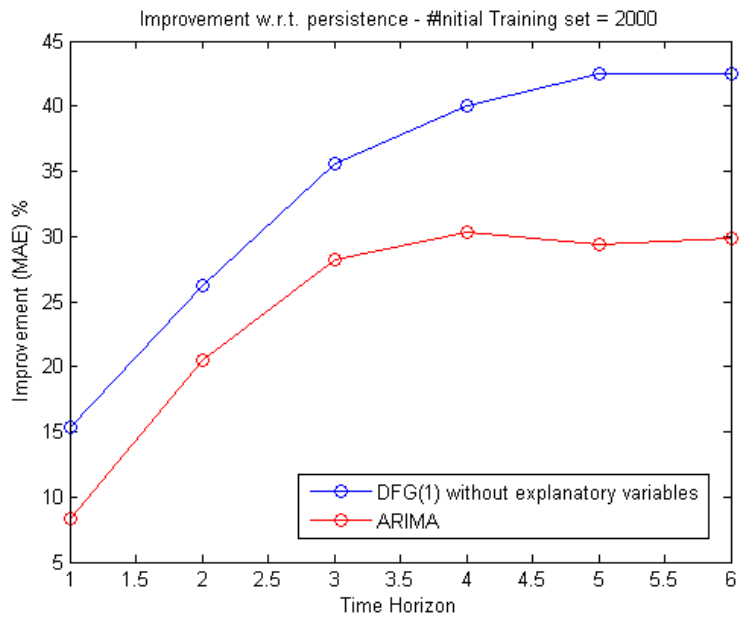


FIGURE 44 WFC for 60-minute time step: MAE improvements w.r.t. persistence for the DFG models with Markov order 1 and ARIMA

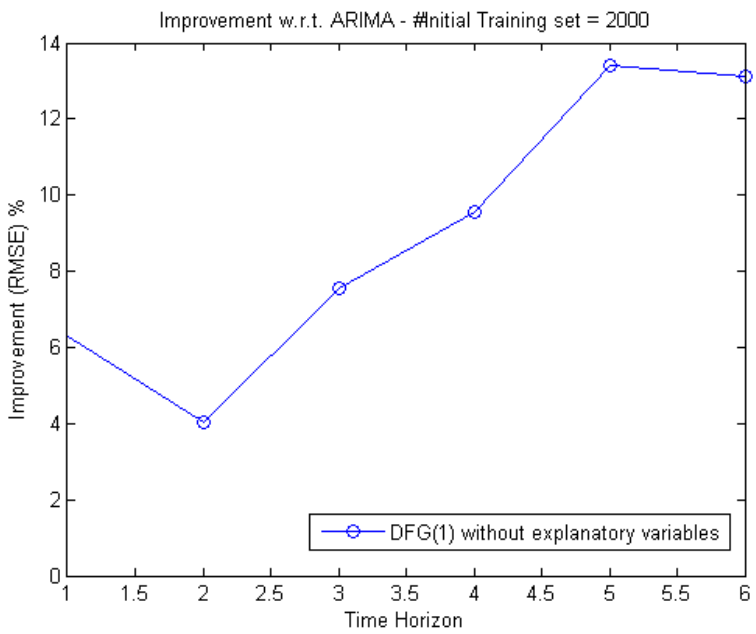


FIGURE 45 WFC for 10-minute time step: RMSE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

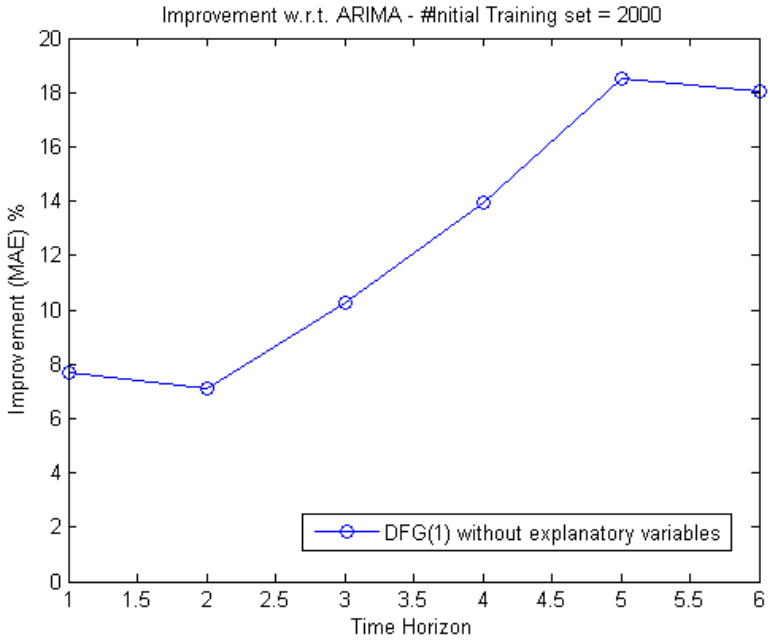


FIGURE 46 WFC for 10-minute time step: MAE improvements w.r.t. ARIMA class for the DFG models with Markov order 1

Overall, we conclude that for the WFC dataset considering the 10, 30, and 60-minute frequencies, the DFG approach performed better than both the persistence and the best ARIMA model benchmarks.

Finally, we did not conduct experiments with NWP explanatory variables for WFC either, again due to the frequent gaps in the available NWP data that made meaningful statistical forecasting experiments and comparisons very challenging. However, the next section describes a different approach we used to analyze the impact of NWPs on the accuracy of the wind power forecasts.

8 WIND TO POWER (W2P) MODEL WITH NWP RESULTS

Due to the lack of continuity in the NWP data and for comparison purposes, we used a wind to power (W2P) model to test the impact of NWP forecasts on forecast performance. The W2P approach does not need the long sequence of continuous data required for time-series-based models. This section presents the results for the very short term wind power point forecasts obtained for the Midwest WFA and a second wind farm in the South Central United States labeled wind farm D (WFD). The experiments were performed by using wind power point forecast methods from the ARGUS-PRIMA (for "prediction intelligent machine") software platform[3] that we developed in previous work. ARGUS-PRIMA is used to test advanced statistical algorithms for short-term wind power forecasting. The platform, which consists of a set of statistical algorithms to generate wind power point and uncertainty forecasts, has been used for systematic testing and comparison of different computational learning algorithms [21]. For wind power point forecasting, ARGUS-PRIMA uses concepts from information theoretic learning (ITL) for training an ANN. A key feature of ITL is that it does not assume a Gaussian probability distribution for the forecasting errors.

In the current study, we developed wind power predictions with 30-minute and 1-hour time resolutions by using the ANN model trained with NWPs (wind speed and direction) and the hour of the day (circular variable) as explanatory inputs. The results were compared to a persistence benchmark model.

Where applicable, the ANN was trained with the maximum correntropy criterion (MCC) criterion, for both offline and online modes. All the W2P models were trained with the classic batch (offline) training, with the maximum number of epochs being the stop training criterion (800 epochs in these tests). The MCC learning rate is set to 0.1, the kernel size σ_{MCC} is 0.1 after the first 200 iterations, and for the remaining 600 epochs, a classic batch backpropagation training algorithm is used. A time-adaptive (online) training algorithm was also tested using a learning rate of 0.001 and momentum factor of 0.0.

8.1 MIDWEST WIND FARM A (WFA)

8.1.1 Input Data

The complete set of input data available for the study discussed here for WFA corresponds to the period between December 1, 2011, and August 31, 2012. The data have two different temporal resolutions: 10 minutes for the SCADA data from the wind farm, and 15 minutes for the NWPs; the forecast time stamps chosen were 30 minutes and 1 hour. Preprocessing and aggregating the data consisted of calculating the 30-minute and 1-hour means, matching the corresponding time stamps, and filtering out the remainder of the data that would not be used in the forecasts.

[3] http://www.anl.gov/technology/project/argus-prima-wind-power-prediction.

Due to the large number of influential parameters, several assumptions were made for the purpose of this report. With regard to the NWP data, there are 14 NWP points distributed geographically over the wind park, from which only one point was chosen. Furthermore, in order to train and evaluate the wind power forecasts, the complete set of available data was split into two datasets: a training data set with targets (desired values) known in advance, and a testing data set for which the realizations were predicted by the trained W2P model. Hence, the following data partition was implemented:

- Training set: December 1, 2011–May 31, 2012, and
- Testing set: June 1, 2012–August 31, 2012.

Wind power forecasts were launched at each hour *T* of the day for the next 15 hours (i.e., *T* + 0–15 hours ahead) and compared to the respective wind power measurements between *T* and *T* + 15 hours. Since the predictions had a 30-minute or 1-hour temporal resolution, for each 15-hour-ahead forecast, there were 31 or 16 predictions, respectively.

8.1.2 Results Performance Analysis

This subsection presents the best results obtained by using the time-adaptive training algorithm with the MCC training criterion compared with the results of the persistence model. Figures 47 through 50 show the normalized mean absolute error (NMAE), normalized root mean square error (NRMSE), normalized bias (NBIAS), and normalized standard deviation error (NSDE), respectively, for different forecast horizons.

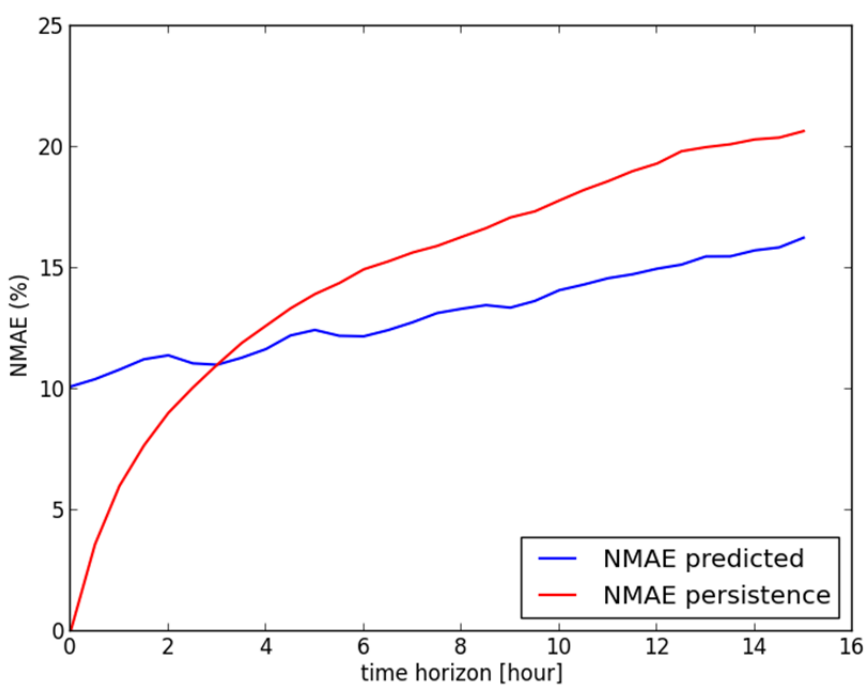


FIGURE 47 WFA: NMAE results comparison

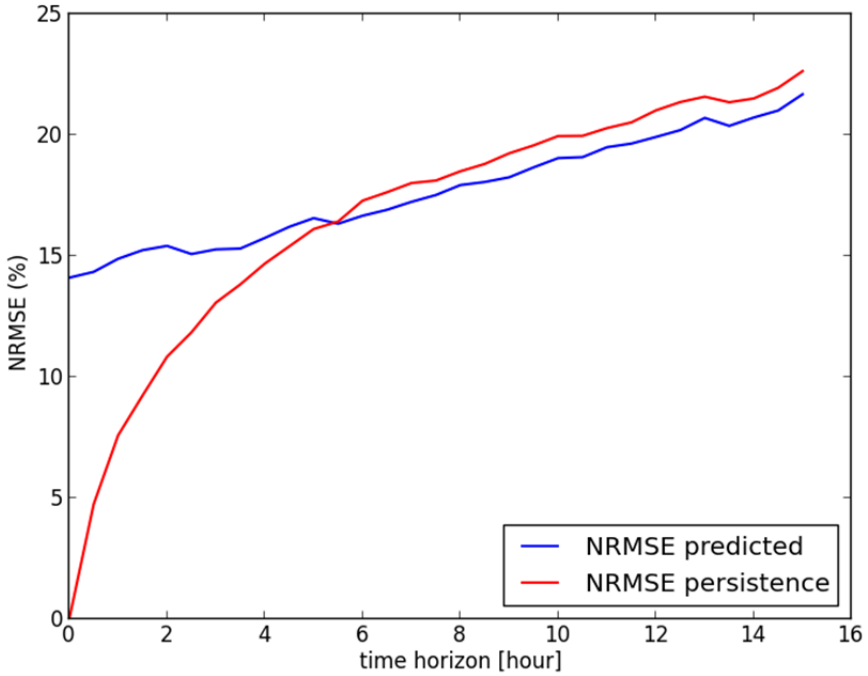


FIGURE 48 WFA: NRMSE results comparison

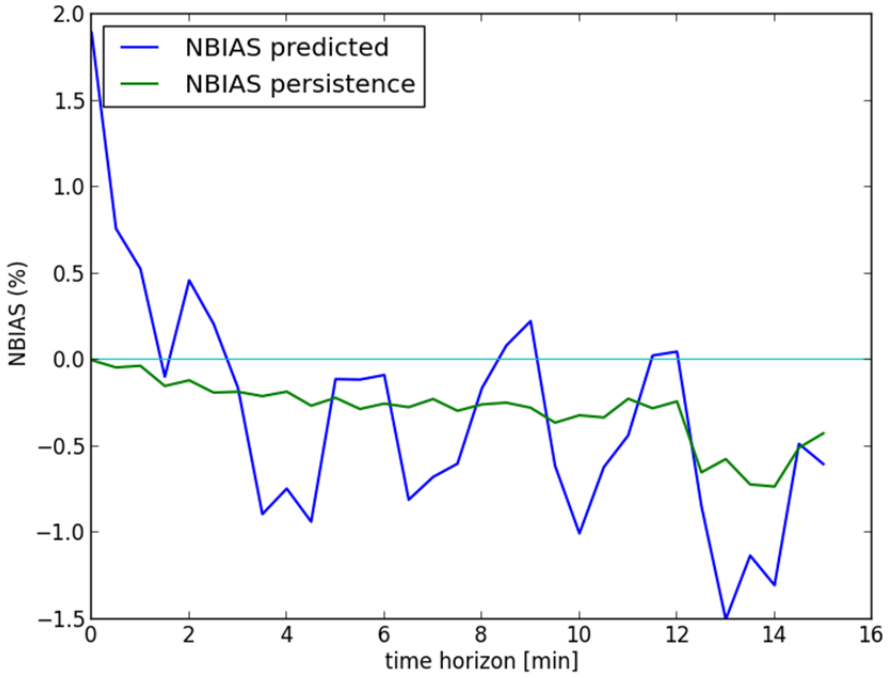


FIGURE 49 WFA: NBIAS results comparison

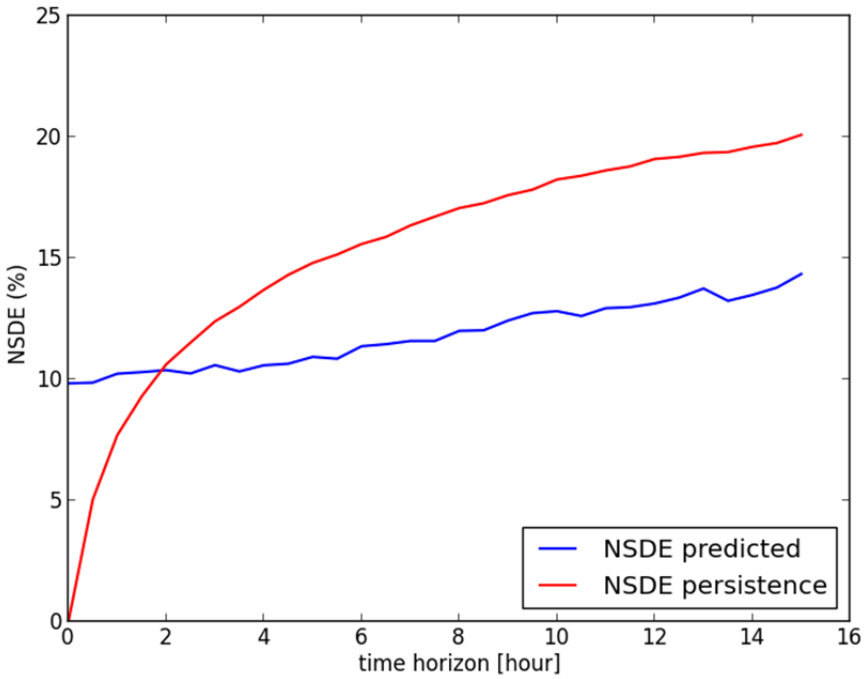


FIGURE 50 WFA: NSDE results comparison

The results presented in Figures 47 through 50 indicate that using NWPs as input to the W2P model of the ARGUS-PRIMA platform in order to make very short term forecasts is not satisfactory with regard to obtaining good wind power point prediction results. In fact, this model can overcome persistence only after the first 3 to 4 hours (NMAE) and 5 to 6 hours (NRMSE). Other types of statistical models, like the DFG algorithms presented in Sections 3 and 4, are clearly better-suited for the 0- to 6-hour time frame.

8.1.3 Evaluation of the Wind Speed NWPs

This subsection presents a study on the quality of the wind speed prediction results from the NWP model, as a complementary means for evaluating the performance of the very short term wind power point forecasts presented above. The wind speed NWPs were averaged over 30-minute data and 1-hour data in order to be used as inputs to the wind power prediction model. Thus, the quality analysis of the NWPs is divided into these two sets. For this study, we considered all the available data used to train and test the W2P model (from December 2011 to August 2012).

30-Minute Data Results

Figure 51 shows the power curve, comparing the dependence of the measured power to the predicted wind speed and the actual measured wind speed. The latter exhibits a typical sigmoid, whereas the first fills almost the whole domain of power and wind speed.

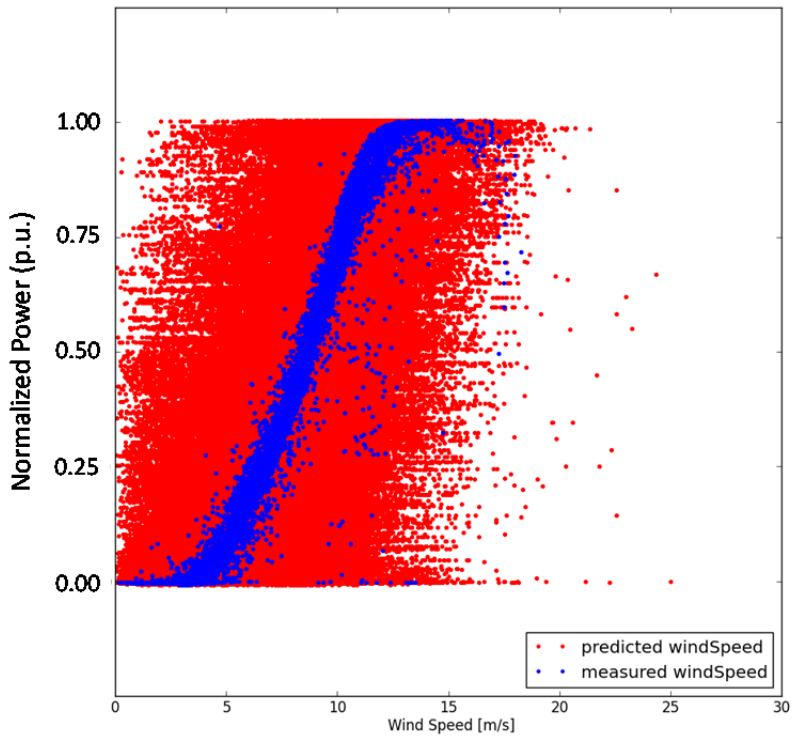


FIGURE 51 WFA: Power curve for 30-minute data with predicted and measured wind speeds

Figures 52 through 57 exhibit the NWP error results for the 30-minute wind speed, normalized to the maximum wind speed value. Figures 52 through 57 show that for 30-minute averaged wind speed data, based on the NMAE and RMSE, the wind speed forecasts become better than the persistence results after 7 hours ahead, when the improvement varies between 0% and 25%. Moreover, the standard deviation of the forecasting errors varies by around 10% throughout the whole time horizon, overcoming the persistence model's at 7 hours ahead (Figure 56). With regard to the NBIAS of the wind speed predictions, it increases during the first 4 hours, then levels out between –3% and –4.5%. In contrast, the bias of the persistence model is small, with limited variation along the horizon (Figure 57).

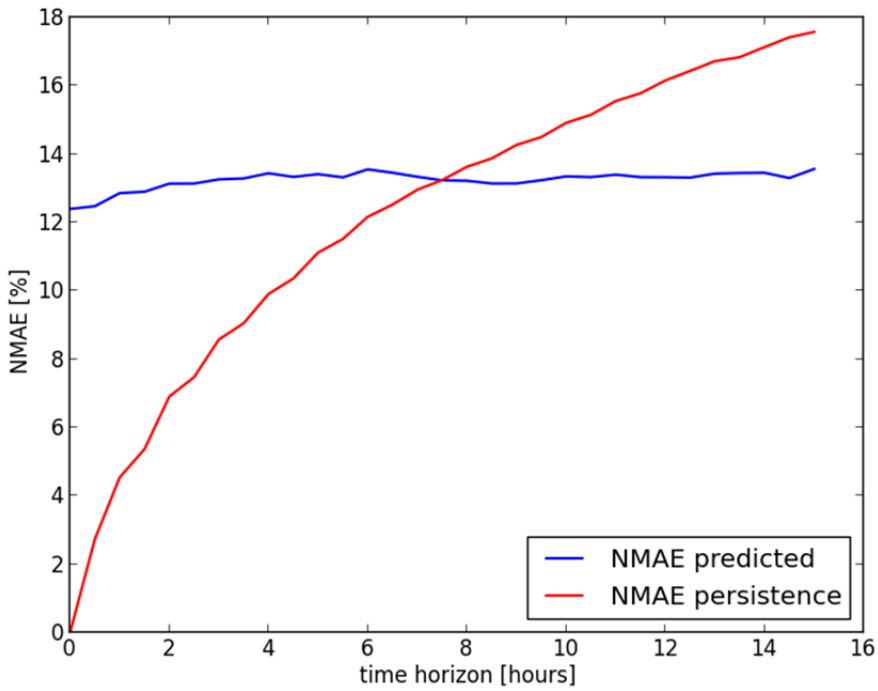


FIGURE 52 WFA: NMAE comparison between NWP wind speed and persistence forecasts for 30-minute data

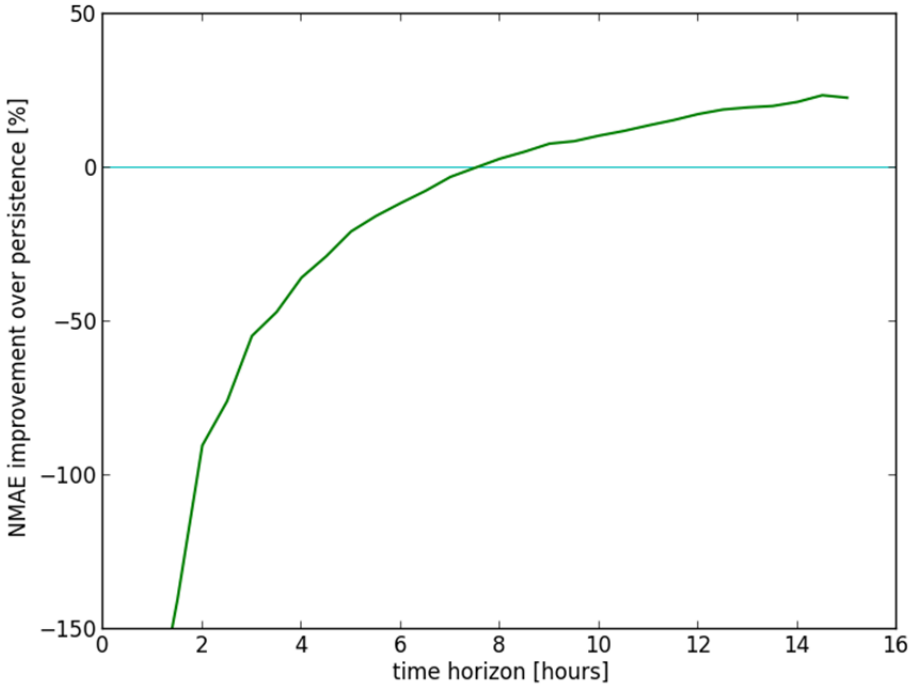


FIGURE 53 WFA: NMAE improvement over persistence for 30-minute data

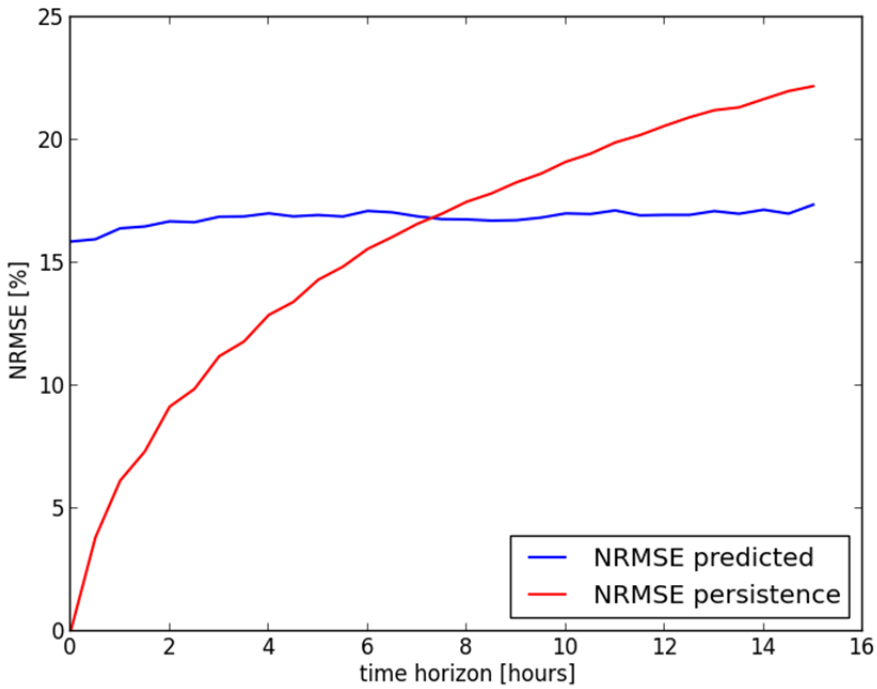


FIGURE 54 WFA: NRMSE comparison between NWP wind speed and persistence forecasts for 30-minute data

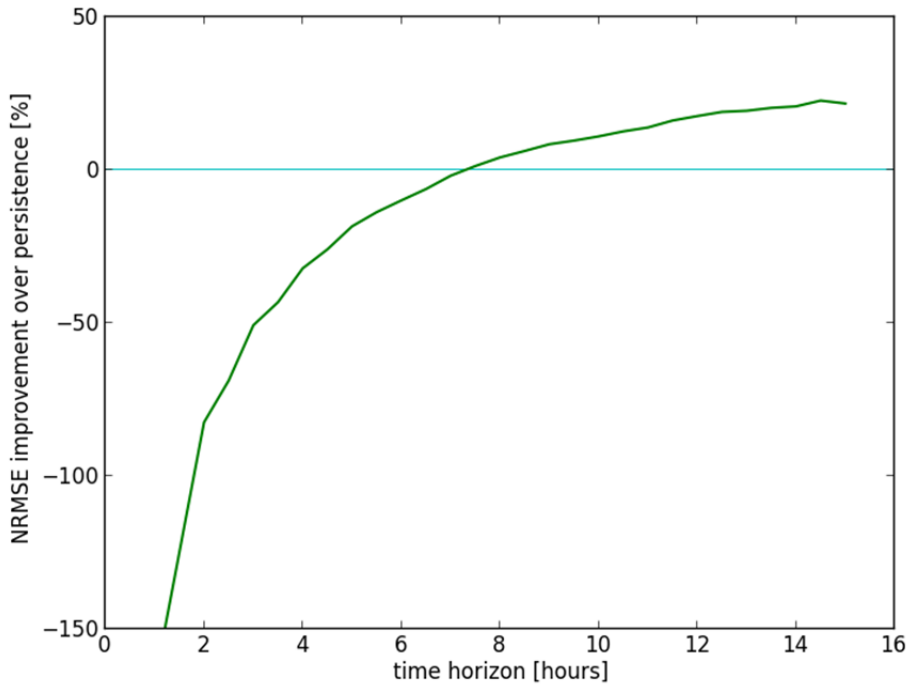


FIGURE 55 WFA: NRMSE improvement over persistence for 30-minute data

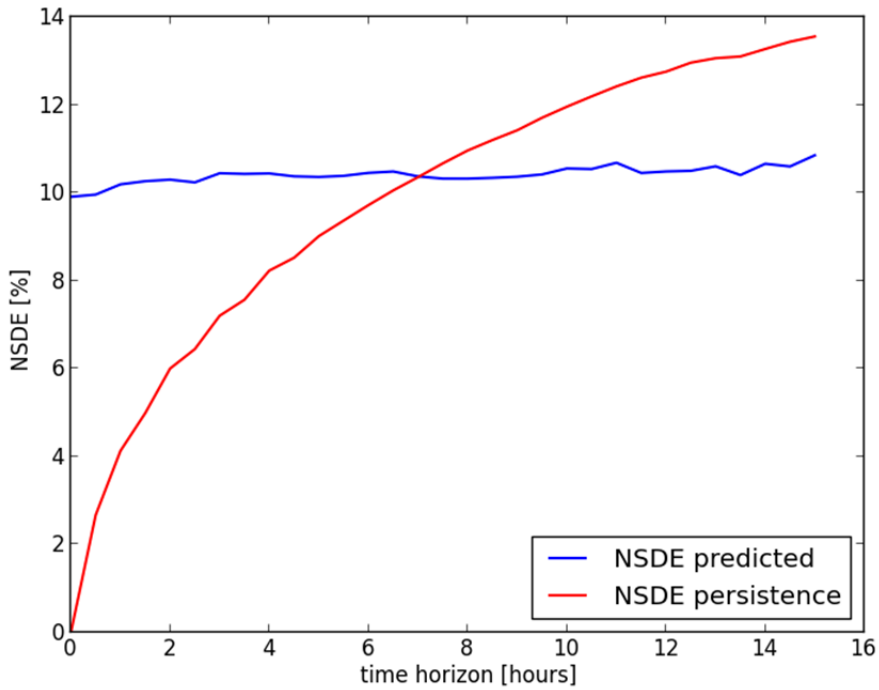


FIGURE 56 WFA: NSDE comparison between NWP wind speed and persistence forecasts for 30-minute data

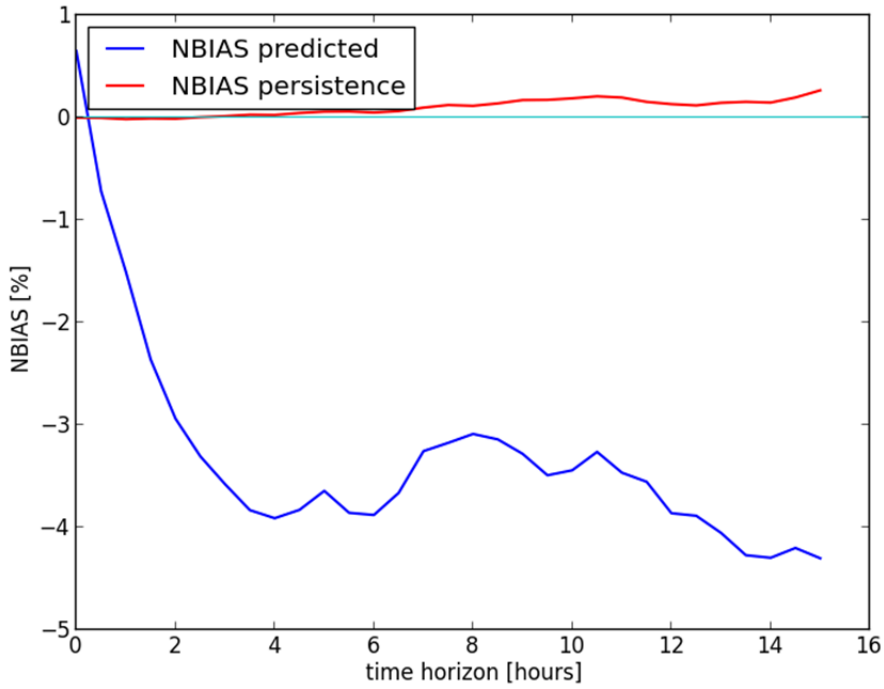


FIGURE 57 WFA: NBIAS comparison between NWP wind speed and persistence forecasts for 30-minute data

1-Hour Data Results

We repeated the analysis by using wind speed data with a 1-hour resolution. Figure 58 shows the power curve, comparing the dependence of the measured power on the predicted and measured wind speed. Again, the latter exhibits a typical sigmoid, whereas the first covers a very large area in the power–wind speed domain. Figures 59 through 64 exhibit the NWP error results for the 1-hour wind speed, normalized to the maximum wind speed value.

Similar to the previous case, for the 1-hour averaged wind speed data, the NMAE, NRMSE, and standard deviation of the errors become better than those of the persistence benchmark after 7 hours ahead. The metrics have only a limited variation, around 12%, 15%, and 10%, respectively. The normalized bias is always higher than the one for persistence, growing with the forecast horizon until –4.8% at 15 hours ahead.

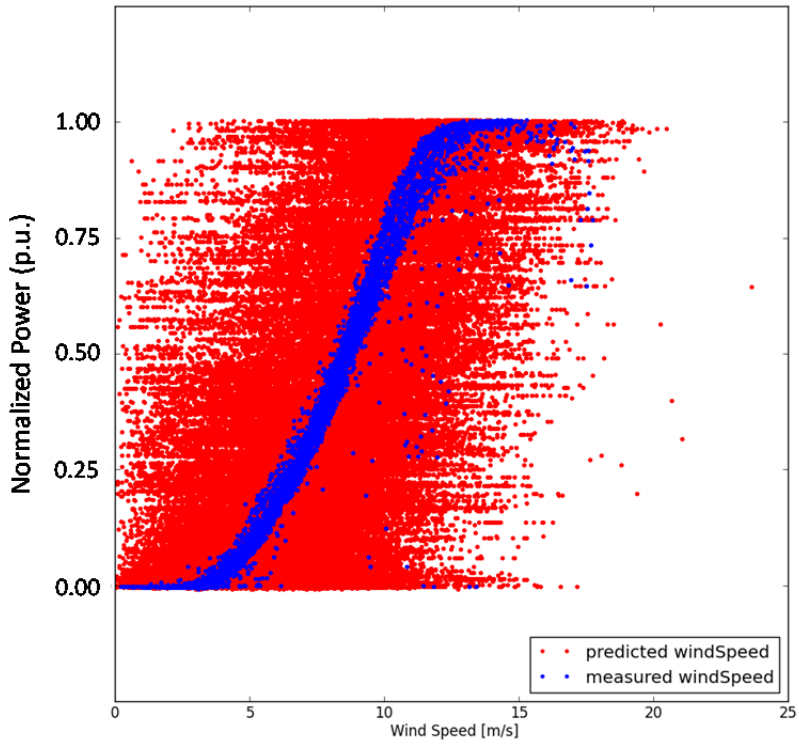


FIGURE 58 WFA: Power curve for 1-hour data

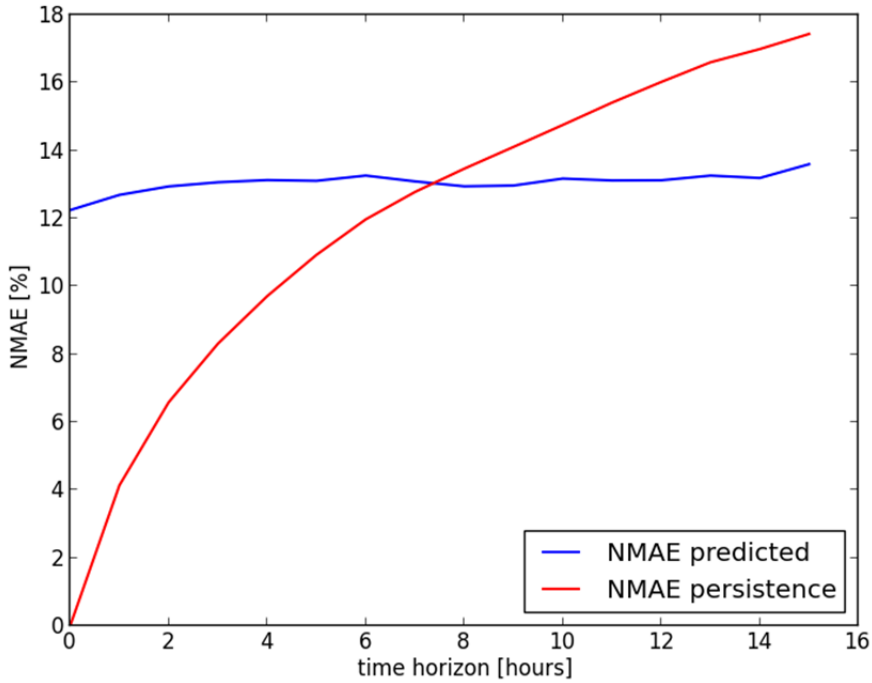


FIGURE 59 WFA: NMAE comparison between NWP wind speed and persistence forecasts for 1-hour data

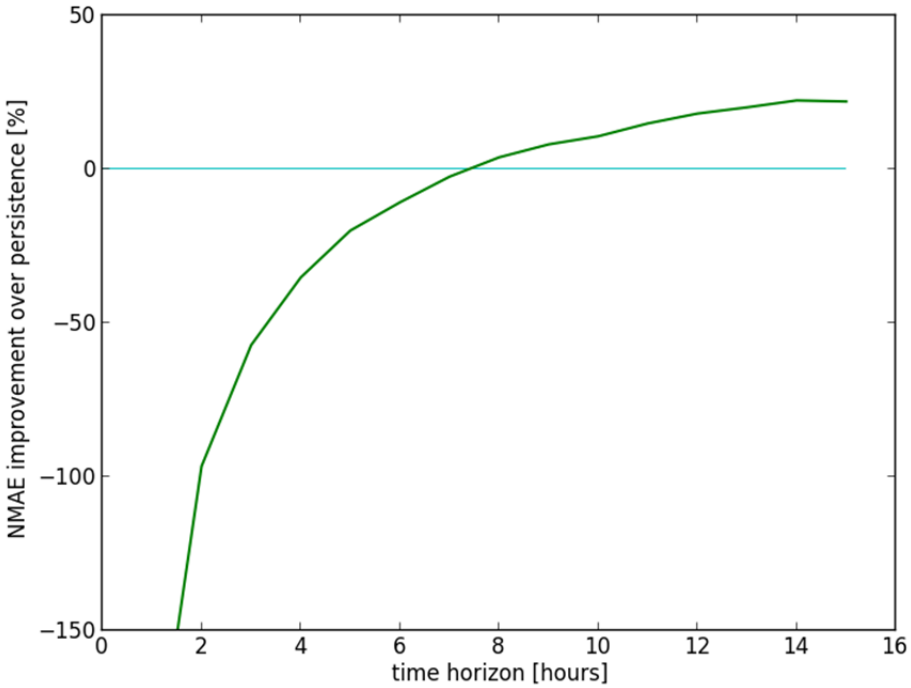


FIGURE 60 WFA: NMAE improvement over persistence for 1-hour data

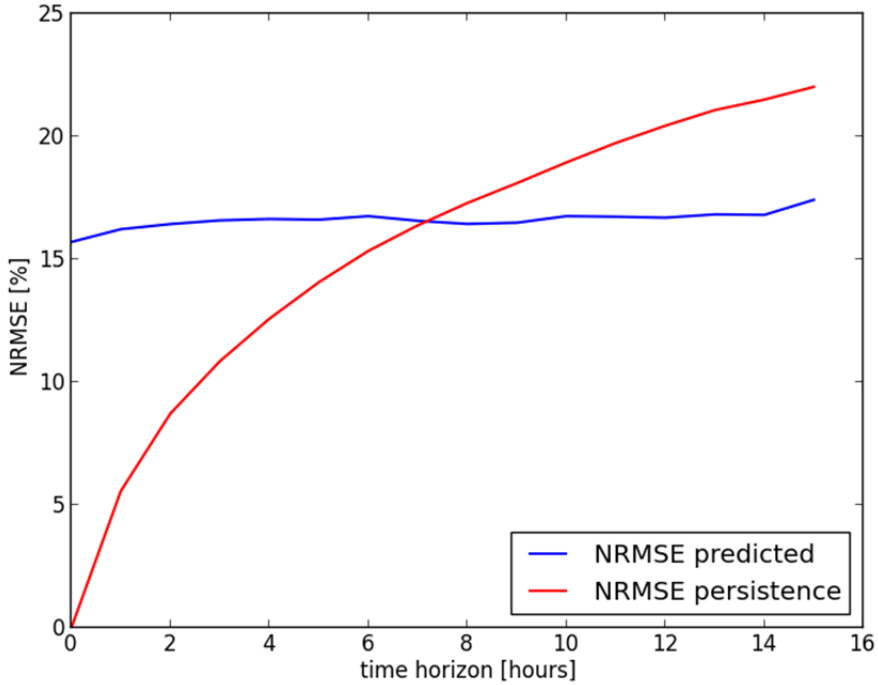


FIGURE 61 WFA: NRMSE comparison between NWP wind speed and persistence forecasts for 1-hour data

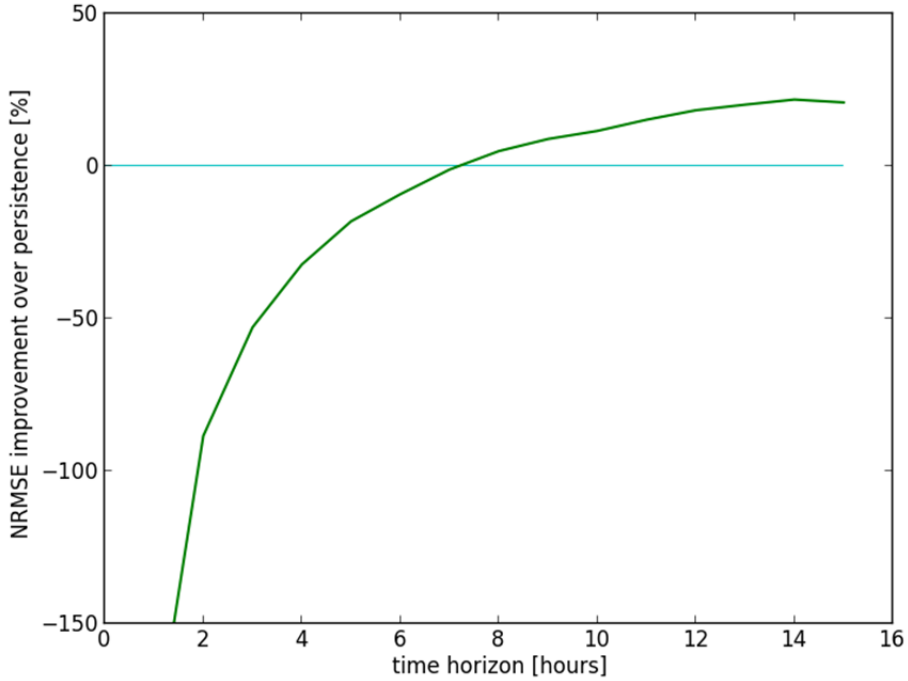


FIGURE 62 WFA: NRMSE improvement over persistence for 1-hour data

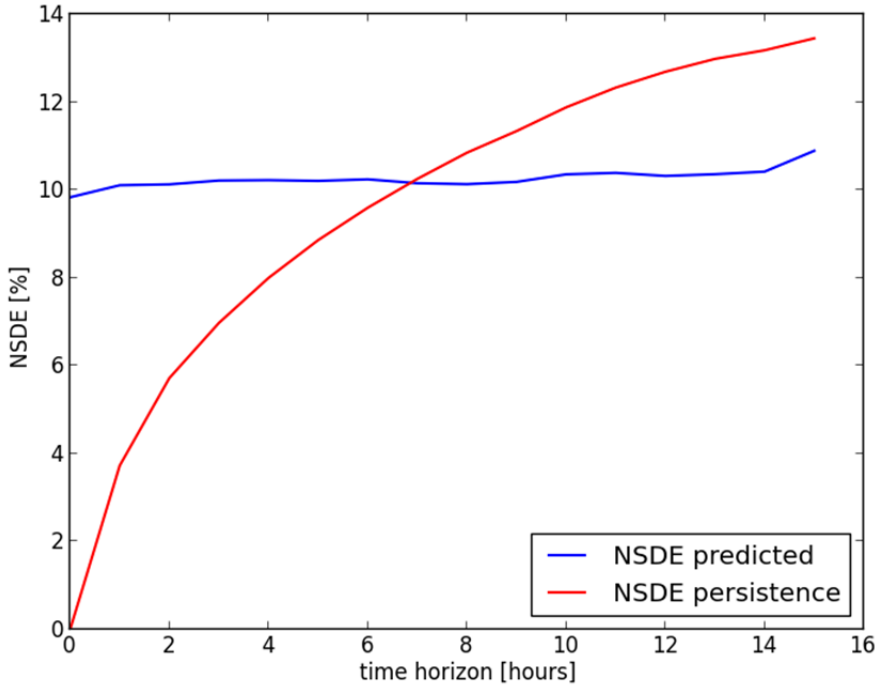


FIGURE 63 WFA: NSDE comparison between NWP wind speed and persistence forecasts for 1-hour data

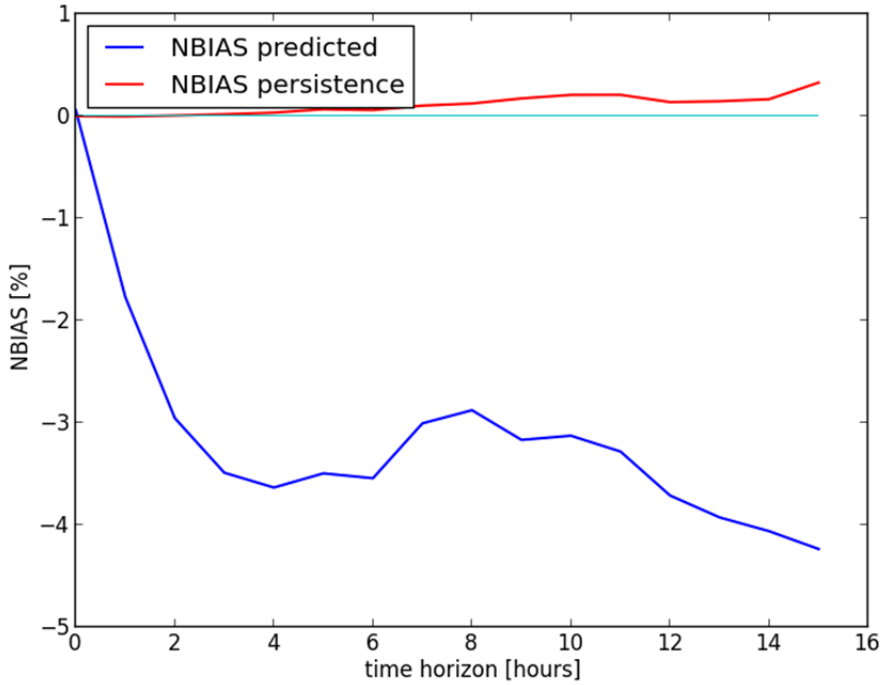


FIGURE 64 WFA: NBIAS comparison between NWP wind speed and persistence forecasts for 1-hour data

8.1.4 Summary of Results

Results for the very short term wind power prediction using the ARGUS-PRIMA W2P point forecasting model for WFA show that the NMAE varies between 10% and 15% and the NRMSE varies between 15% and 20%. Moreover, online training improves the forecast performance. However, the W2P model surpasses the accuracy of the persistence model only after 3 hours (NMAE) and 5 hours (NRMSE).

The analysis of the quality of the NWP wind speeds for the WFA location revealed they are not as accurate as expected. The errors lay between 12% and 16% (NMAE and NRMSE), and the NWP wind speed forecasts can beat the persistence forecast only for forecast horizons of 7 hours or more. Because NWPs are the main inputs of the wind power prediction model used in this section, it can be concluded that the relatively low accuracy of the NWP has compromised the performance of the very short term wind power forecasts. Therefore, although the adaptation of ARGUS-PRIMA to the very short term is not a satisfactory approach for obtaining good wind power point predictions, particularly for time horizons of less than 3 hours, the algorithm's performance could be improved upon the retrieval of better NWPs.

8.2 SOUTH CENTRAL WIND FARM D (WFD)

8.2.1 Input Data

The complete input dataset available for this study corresponds to the period between December 1, 2011, and March 17, 2012. It was delivered in two different temporal resolutions: 10 minutes for the SCADA data, and 15 minutes for the NWPs. The forecast time stamps chosen were 30 minutes. Preprocessing and aggregating the data consisted of calculating the 30-minute means, matching the corresponding time stamps, and filtering out the remainder of the data that would not be used in the forecasts.

Due to the large number of influential parameters, several assumptions were made for the purpose of this report. For the NWP data, there were 9 NWP points distributed geographically over the wind park, from which only 1 point was chosen.

In order to train and evaluate wind power forecasts, the complete set of available data was split into two data sets: a training data set with targets (desired values) known in advance, and a testing data set where the realizations are predicted by the trained W2P model by using the ARGUS-PRIMA platform. Hence, the following data partition was implemented:

- Training set: December 1, 2011–February 29, 2012, and
- Testing set: March 1, 2012–March 17, 2012.

Wind power forecasts were launched at each hour T of the day for the next 15 hours (i.e., T + 0–15 hours ahead) and compared to the respective wind power measurements between

T and *T* + 15 hours. Since the predictions have a 30-minute temporal resolution, for each launch time, there were 31 predictions.

This subsection presents the results for the very short term wind power point forecasts obtained for WFD located in the South Central United States. The point forecasts were performed by using the ARGUS-PRIMA platform in two different configurations: an ANN trained with NWP inputs (wind speed and direction), and an ANN trained with the same NWP inputs but that also considered the last-known realization of wind power as an additional input. The study made 30-minute step predictions, all of which were compared, with persistence as the benchmark model.

Several tests were performed in order to identify the best ANN. Different neural networks were trained with MCC, mean square error (MSE), and combined MSE-MCC criteria. All these models were trained with the classic batch (offline) and the adaptive (online) training approaches, with the maximum number of epochs being the stop training criterion. The number of epochs was variable and treated as a parameter, as listed in Section 8.2.2. Several values of the MCC learning rate and the kernel size σ_{MCC} were also considered, in order to find the best weights for the ANN. The specific values for the different experiments are also listed in Section 8.2.2.

8.2.2 Results Performance Analysis

This subsection presents a comparison of performance between the wind power point predictions and persistence forecasts for the four best model configurations.

Using Only the NWPs as Input

The best ANN training criteria and the corresponding parameters for these inputs were found to be as follows:

- R1: 3,000 epochs MCC, learning rate = 0.05, $\sigma_{MCC} = 0.01$, and
- R2: 3,000 epochs MSE.

Figures 65, 66, and 67 show NMAE, NRMSE, and NBIAS, respectively, for results R1 and R2.

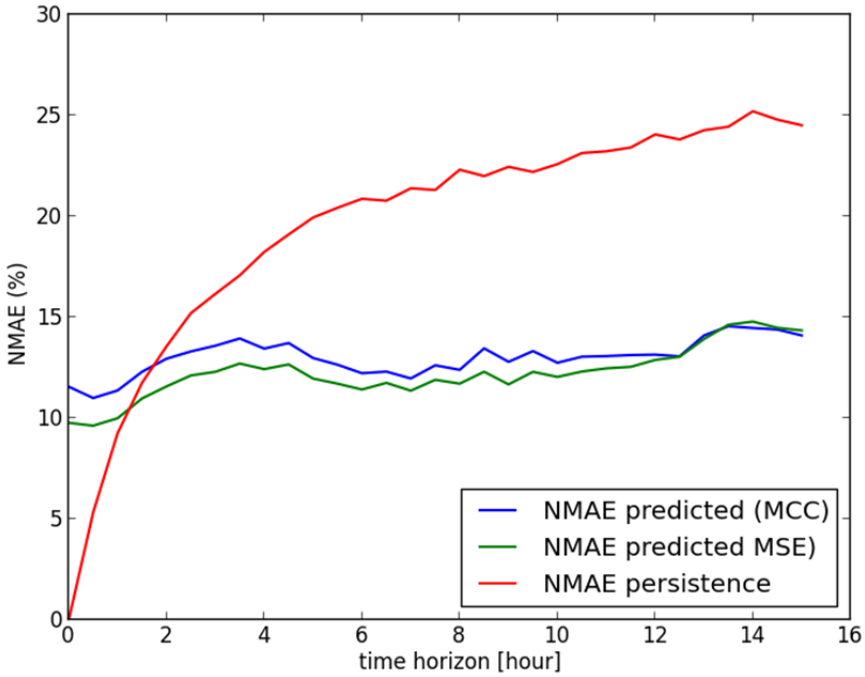


FIGURE 65 WFD: NMAE comparison between the two prediction models (R1, R2) and persistence

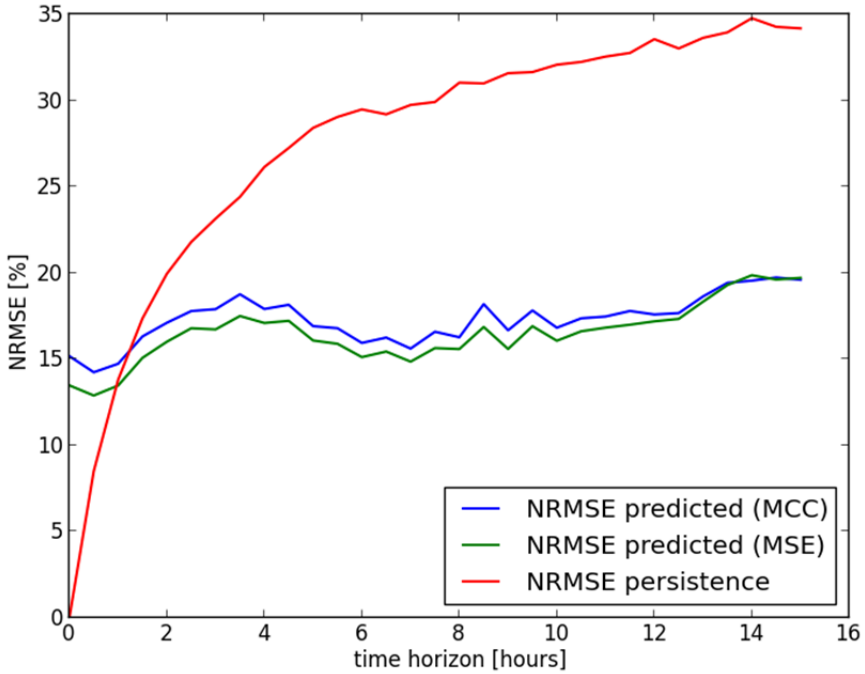


FIGURE 66 WFD: NRMSE comparison between the two prediction models (R1, R2) and persistence

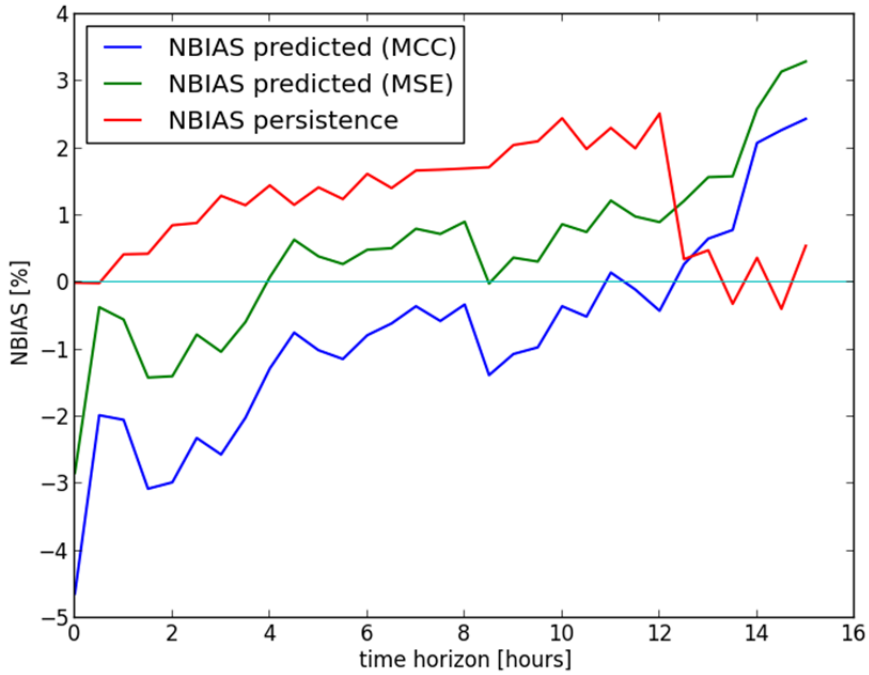


FIGURE 67 WFD: NBIAS comparison between the two prediction models (R1, R2) and persistence

The improvements over the persistence model are presented in Figures 68 and 69. From the two graphs, it can be seen that the MSE training criterion (R2) performs better than the MCC training criterion (R1). In addition, the results show that the W2P model performs better than the persistence model after the first 1 or 2 hours for both training criteria.

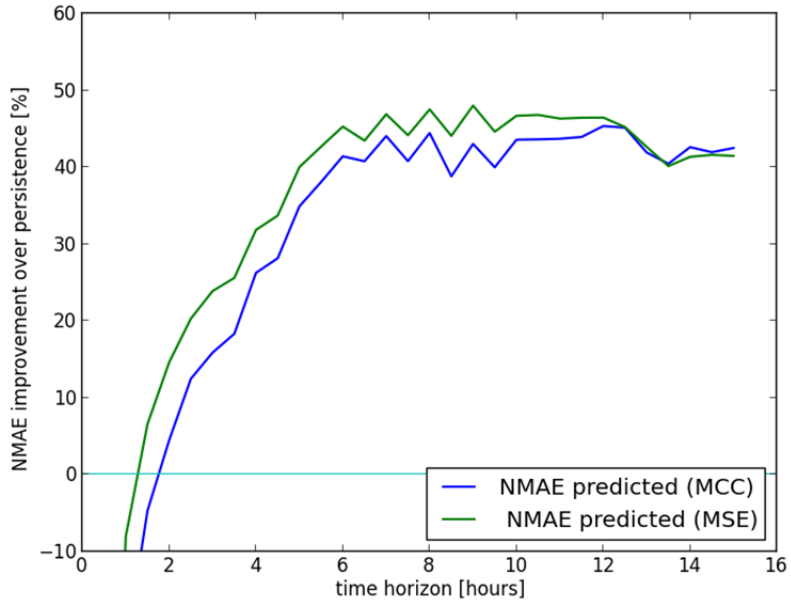


FIGURE 68 WFD: NMAE improvement over persistence for the two prediction models (R1, R2)

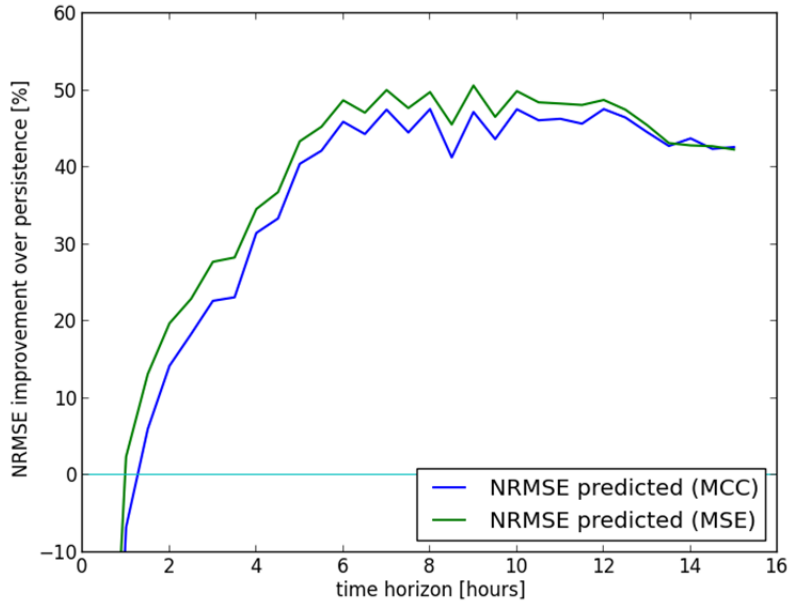


FIGURE 69 WFD: NRMSE improvement over persistence for the two prediction models (R1, R2)

Using the Last-Known Realization as an Additional Input

In the analysis presented next, we used the last known realization of wind power as an input in addition to the NWP inputs, with the following ANN parameters:

- R3: 3,000 epochs MCC, learning rate = 0.05, $\sigma_{MCC} = 0.1$, considering also the last-known realization as an input, and
- R4: 3,000 epochs MSE, considering the last-known realization as an additional input.

Figures 70, 71, and 72 show the NMAE, NRMSE, and NBIAS, respectively, for results R3 and R4, whereas Figures 73 and 74 show the relative performance compared to the persistence benchmark.

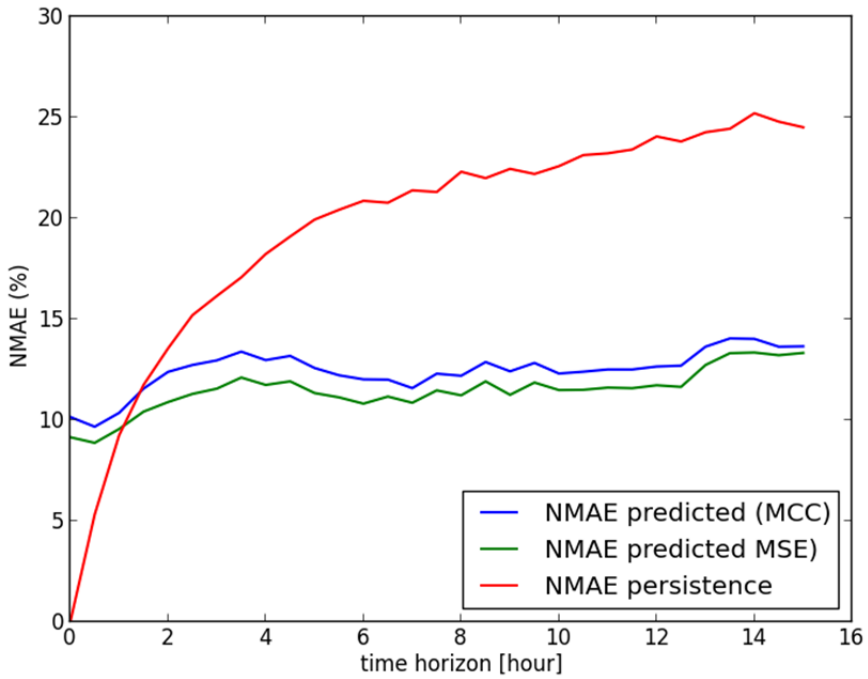


FIGURE 70 WFD: NMAE comparison between the two prediction models (R3, R4) and persistence

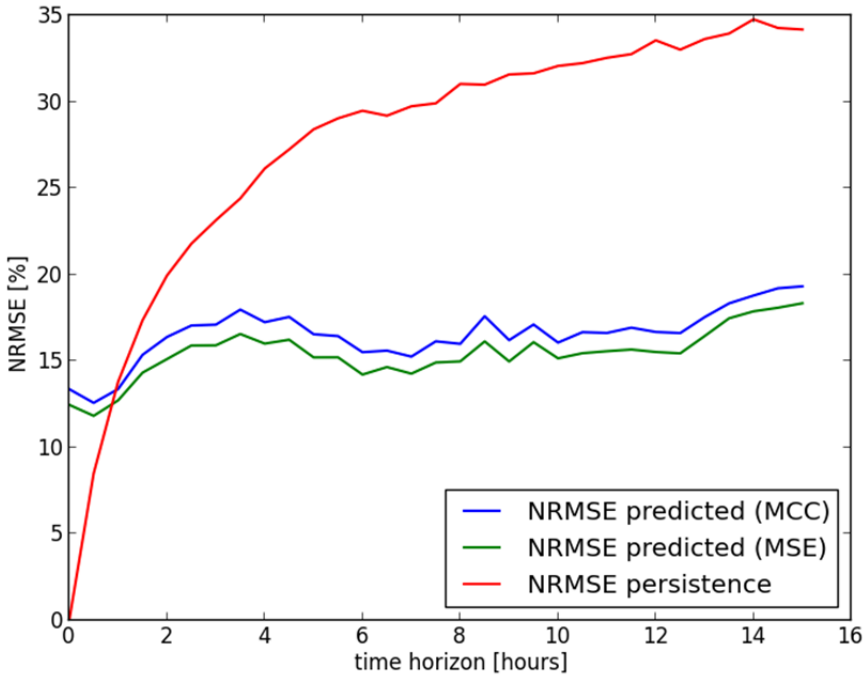


FIGURE 71 WFD: NRMSE comparison between the two prediction models (R3, R4) and persistence

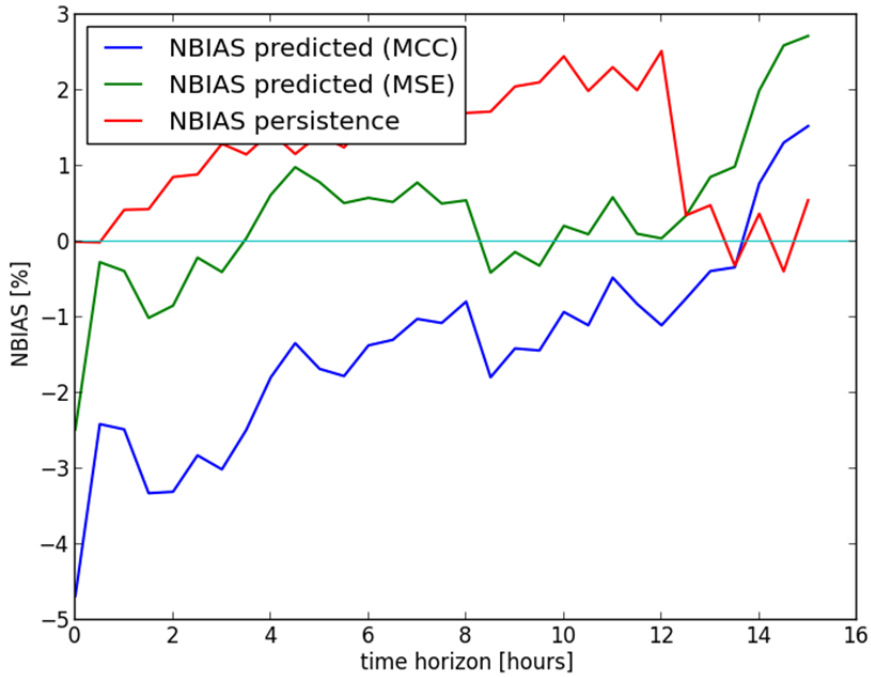


FIGURE 72 WFD: NBIAS comparison between the two prediction models (R3, R4) and persistence

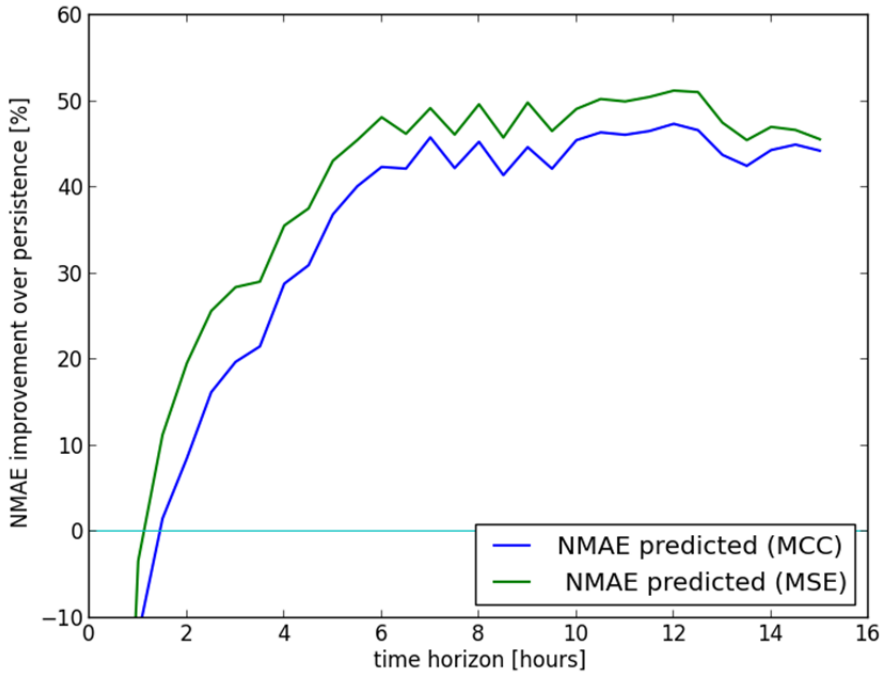


FIGURE 73 WFD: NMAE improvement over persistence for the two prediction models (R3, R4)

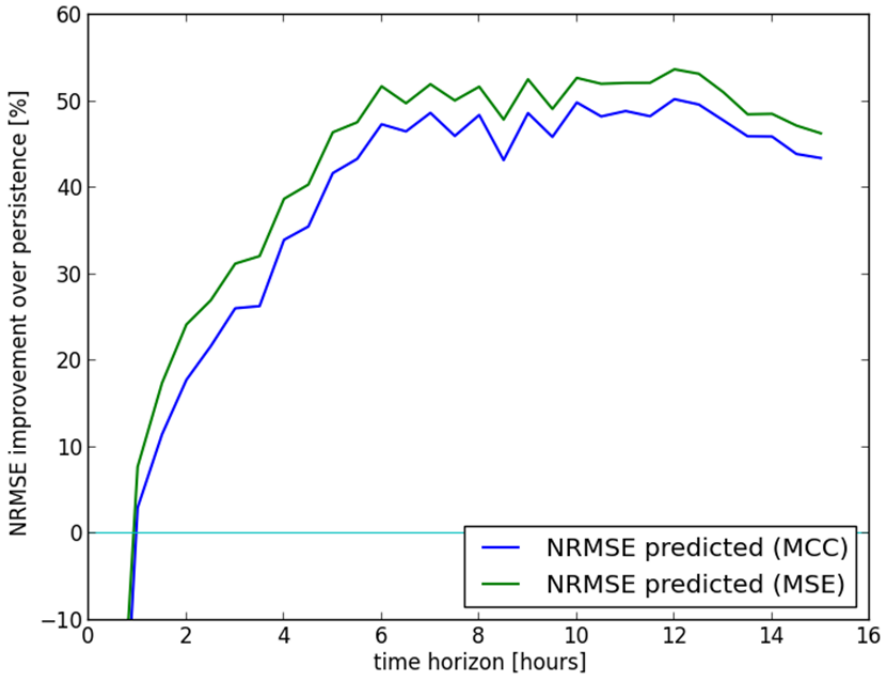


FIGURE 74 WFD: NRMSE improvement over persistence for the two prediction models (R3, R4)

From the results presented in Figures 70, 71, and 72, note again that the MSE training criterion (R4) performs better than the MCC training criterion (R3). The NMAE varies between 10% and 15%, and the NRMSE varies between 13% and 20% in the case of the pure W2P models. In the other case, with the use of the last-known wind power realization, the NMAE varies between 9% and 14%, and the NRMSE varies between 12% and 20%. Hence, the addition of the last- known realization as an input improves the performance of the W2P model. Again, the W2P model performs better than the persistence model after the first 1 to 2 hours for both training criteria.

8.2.3 Summary of Results

Results for the very short term wind power prediction using the ARGUS-PRIMA model for WFD show that, depending on the accuracy of the NWPs, a W2P model can perform well for very short term forecasting. However, it is very hard to beat the persistence model for very short horizons (i.e., less than 1 to 2 hours). The time-series-based models have better performance at the very short forecast horizon, as shown in Sections 5 to 7. Finally, the analysis in this section shows that the introduction of the last-known wind power realization improves the W2P model's performance, so it overcomes the persistence model for shorter forecast time horizons.

9 CONCLUDING REMARKS

This report demonstrates the capabilities of DFGs in modeling linear relationships between wind power and both its lagged values and external variables. The DFGs are shown to be a reliable tool for wind power forecasting. In particular, they outperform the class of ARIMA models because they are more flexible in describing time patterns. That added flexibility seems to pay off for the WPF application.

However, the DFG approach has some limitations, particularly in computational terms. The processing time increases cubically with the number of points (N) in the training set; that is, the computational complexity of DFG models is of the order $O(N^3)$. For the experiments presented in Sections 5, 6, and 7, the algorithms required an average computational time of around 8 to 10 hours for N = 5,000 and 6 to 8 hours for N = 2,000. The DFG model structure implies that each time the training sample is enlarged by the addition of the previous test dataset, the entire *Z* sequence must be re-estimated. The associated computational load quickly becomes severe. We tried to bypass this limitation by developing the new data training approach presented in Subsection 4.1, which is based on the idea of fixing the estimated *Z* sequence from the initial training sample and estimating separately the new values over the previous test set.

The experimental results in this report show the potential gain from using the DFG class of models for forecasting wind power generation. Our empirical findings indicate that there is a low level of predictability in wind speed (and therefore wind power). In other words, although wind speed is not a Martingale (the persistence model was outperformed even when the ARIMA approach was used), wind speed does seem to be fairly volatile and display a low level of memory. This report shows the results of applying first Markov order DFG models to three wind farms and also making use of NWP information when available. It concentrates on very short term forecasting horizons (i.e., those up to 6 hours ahead).

From the results presented in Sections 5, 6, and 7, we can conclude that the DFG models were adequate, particularly for the datasets recorded at a 1-hour sampling frequency. The average RMSE improvements achieved by the DFG models w.r.t. the ARIMA class were 5.18% for the WFA dataset, 13.94% for the WFB dataset, and 9.00% for the WFC dataset. The corresponding average MAE improvements were 7.62%, 15.94%, and 12.60%, respectively.

For the higher sampling frequency (10-minute time step), the DFG approach was also able to outperform the ARIMA models, except for the shorter time horizons. The average RMSE improvements achieved by the DFG models w.r.t. the ARIMA class were 1.79% for the WFA dataset, 0.88% for the WFB dataset, and 1.31% for the WFC dataset. The corresponding average MAE improvements were 3.82%, 2.51%, and 3.51%, respectively.

For the 30-minute sampling frequency, the DFG approach outperformed the ARIMA models much more clearly than it did for the 10-minute sampling frequency. The average RMSE improvements achieved by the DFG models w.r.t. the ARIMA class were 1.64% for the WFA dataset, 2.54% for the WFB dataset, and 3.14% for the WFC dataset. The corresponding average MAE improvements were 3.84%, 6.20%, and 5.69%, respectively.

It was observed that the DFG forecasts tend to smooth out some of the variability in the time series. Hourly data show less variability than their higher-frequency counterparts, making hourly observations more predictable. This advantage of DFG models over ARIMA models confirms that the former are more flexible than the latter.

Weather forecasts from a NWP model (i.e., the HRRR predictions developed for the WFIP) were also combined with past values from the wind power time series within the DFG framework for one wind farm (WFA). This resulted in a further increase in the improvement over the persistence model when compared to a DFG without NWP: an increase of around 4% in MAE and 10% in RMSE for the first lead time. This means that combining past values from the time series of actual wind power with NWP predictions can result in an improved power forecast. Nevertheless, for longer time horizons (i.e., more than two lead times), this improvement is marginal according to our results.

Within the DFG framework, we also studied the possibility of exchanging the quadratic energy function with a cost function derived from ITL principles: the correntropy. We successfully applied this alternative cost function, but it did not lead to significantly better forecasting results. For most datasets, we obtained results similar to those derived from the quadratic energy cost function. For this reason, we did not report the results based on the alternative correntropy function in this report.

Finally, we tested the quality of the NWPs by using them in a W2P model for WFA and WFD. We concluded that the addition of NWP inputs, at the level of quality available for this study, could not, by itself, improve the accuracy of the predictions over the level offered by a persistence model, for the first forecasting hours, particularly in the case of WFA. However, for WFD, the W2P approach with NWP did provide more accurate forecast results than those of the persistence model for forecast horizons beyond 1 to 2 hours. Also, from the experiments with WFA, we could affirm that even though the NWPs were not good enough to provide, by themselves, accurate predictions of very short term wind power with the W2P approach, NWPs may be used to increase the improvement offered by time-series-based models.

9.1 FUTURE RESEARCH

The DFG class of models allows for many different generalizations. In this study we limited the scope to the use of linear and dynamic functions. Thus, there are many different directions in which this study can be expanded. One relevant direction that should be explored would be to use nonlinear f and h functions. For instance, both of these functions could take the form of a nonparametric interpolator, such as an ANN. This extension has the potential to improve the forecast accuracy of DFG models and represents an important direction for future work. Another direction for further research would be to decrease the computational effort of the DFG approach. One possibility for improving the running time would be to develop new and efficient backpropagation algorithms for this type of model and/or to adopt recursive (or online) optimization techniques. Finally, testing should be conducted on more wind farms and by using different NWPs to confirm the findings in this study.

10 REFERENCES

[1] Monteiro, C., R. Bessa, V. Miranda, A. Botterud, J. Wang and G. Conzelmann, 2009, *Wind Power Forecasting: State-of-the-Art 2009*. Report, ANL/DIS-10-1, Argonne National Laboratory, Argonne, Ill.

[2] Fabbri, A., T.G.S. Roman, J.R. Abbad and V.H.M. Quezada, 2005, "Assessment of the Cost Associated with Wind Generation Prediction Errors in Liberalized Electricity Market," *IEEE Transactions on Power Systems* 20(3):1440–1446.

[3] Spencer, R., 2002, "Untapped Potential of Wind Power," *IEEE Power Engineering Review* 22(9):10–11, Sept.

[4] Chompoo-inwai, C., W. Lee, P. Fuangfoo, M. Williams, and J. Liao, "System Impact Study for the Interconnection of Wind Generation and Utility System," *IEEE Transactions on Industry Application,* vol. 41 (1), pp. 163-168, January/February 2005.

[5] Piwko, R., D. Osborn, R. Gramlich, G. Jordan, D. Hawkins, and K. Porter, 2005, "Wind Energy Delivery Issues," *IEEE Power & Energy Magazine* 3(6):47–56, Nov./Dec.

[6] Ernst, B., B. Oakleaf, M.L. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken, and K. Rohrig, "Predicting the Wind — Models and Methods of Wind Forecasting for Utility Operations Planning," *IEEE Power & Energy* 5(6):78–89, Nov./Dec.

[7] Foley, A., P. Leahy, A. Marvuglia, and E. Mckeogh, 2012, "Current Methods and Advances in Forecasting of Wind Power Generation," *Renewable Energy* 37:1–8.

[8] Potter, C., and M. Negnevitsky, 2006, "Very Short-Time Wind Forecasting for Tasmanian Power Generation," *IEEE Transactions on Power Systems* 21(2):965–972.

[9] Bracale, A., G. Carpinelli, M. Mangoni, and D. Proto, 2009, "Wind Power Forecast Methods and Very Short-Term Steady-State Analysis of an Electrical Distribution System," *Electrical Engineering Research Report* 1(1):1-9.

[10] Mendes, J., J. Sumaili, R. Bessa, H. Keko, V. Miranda, A. Botterud and Z. Zhi, 2013, Very Short-Term Wind Power Forecasting: State-of-the-Art, Report ANL/DIS-14/6, Argonne National Laboratory, Argonne, Ill.

[11] Durbin, J., and S. Koopman, 2012, *Time Series Analysis by State Space Methods,* 2nd Edition, Oxford Statistical Science, Oxford, U.K.

[12] Hamilton, J., 1994, *Time Series Analysis,* Princeton University Press, N.J.

[13] Mirowski, P., 2011, "Time Series Modeling with Hidden Variables and Gradient-Based Algorithms," Ph.D. thesis, Courant Institute of Mathematical Sciences.

[14] LeCun, Y., S. Chopra, R. Hadsell, and F. Huang, 2006, *Predicting Structured Outputs,* MIT Press, Mass.

[15] LeCun, Y., S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, 2006, "A Tutorial on Energy-Based Learning," in *Predicting Strutured Data*, MIT Press, Mass.

[16] Rumelhart, D., G. Hinton, and R. Williams, 2006, "Learning Representations by Back-propagating Errors," 2006, *Nature,* No. 323, pp. 533–536.

[17] Dempster, A.P., N.M. Laird, and D.B. Rdin, 1977, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)* 39(1):1--38.

[18] Ghahramani, Z., 1998, "Learning Dynamic Bayesian Networks," *Lecture Notes in Artificial Intelligence,* pp. 168–197.

[19] Mackay, D.J.C., 2003, *Information Theory, Inference and Learning Algorithms,* Cambridge University Press, U.K.

[20] Hyndman, R., and Y. Khandakar, 2008, "Automatic Time Series Forecasting: The Forecast Package for R," *Journal of Statistical Software* 27(3).

[21] Mendes, J., R.J. Bessa, H. Keko, J. Sumaili, V. Miranda, C. Ferreira, J. Gama, A. Botterud, Z. Zhou, and J. Wang, 2011, *Development and Testing of Improved Statistical Wind Power Forecasting Methods*. Report ANL/DIS-11-7, Argonne National Laboratory, Argonne, Ill.

APPENDIX A: MAXIMUM A POSTERIORI (MAP) APPROXIMATION

One definition of the "marginal" energy of the observed sequence Y that fits nicely into previous Eq. 9 is:

$$E(Y;X,W) = -\frac{1}{\beta} log \int_{D_Z} e^{-\beta E(Y,Z;X,W)} dZ$$

Eq. 33

$$e^{-\beta E(Y;X,W)} = \int_{D_Z} e^{-\beta E(Y,Z;X,W)} dZ$$

Eq. 34

$$P(Y|X,W) = \frac{e^{-\beta E(Y;X,W)}}{\int_{D_Y} e^{-\beta E(Y';X,W)} dY'}$$

Eq. 35

By the Bayes rule $P(Y,Z|X,W) = P(Z|X,Y,W)P(Y|X,W)$, maximizing $P(Y,Z|X,W)$ w.r.t. to Z is akin to maximizing $P(Z|X,Y,W)$ w.r.t. Z since $P(Y|X,W)$ does not depend on Z. Then, using Eq. 6, $arg\ max_Z\ P(Z|X,Y,W) = arg\ max_Z\ e^{-\beta E(Y,Z;X,W)}$. Hence:

$$arg\ \underset{Z}{min}\ E(Y,Z;X,W) = arg\ \underset{Z}{max}\ P(Z|X,Y,W)$$

Eq. 36

Now, to prove that $E(Y;X,W)$ can be approximated by $arg\ min_Z\ E(Y,Z;X,W)$, consider the limit in β of the Eq. 13. Assuming that energy $E(Y,Z;X,W)$ is positive and admits a zero minimum Z_0:

$$\lim_{\beta\to\infty} -\frac{1}{\beta} log \int_{D_Z} e^{-\beta E(Y,Z;X,W)} dZ = \lim_{\beta\to\infty} -\frac{1}{\beta} log \int_{D_Z} \delta_{Z=Z_0}\ e^{-\beta E(Y,Z;X,W)} dZ$$

$$= \lim_{\beta\to\infty} -\frac{1}{\beta} log e^{-\beta E(Y,Z_0;X,W)}$$

$$= E(Y,Z_0;X,W)$$

As pointed out in reference [2], for an observed sequence Y and given a model parameterized by W, the result of the latent variable inference is the minimum energy state of the model for that sequence:

$$E(Y;X,W) = arg\,\min_{Z} E(Y,Z;X,W)$$

Eq. 37

APPENDIX B: DERIVATION OF THE GRADIENT FUNCTION

The component of the gradient corresponding to the initializations of the latent variable can be written as:

$$\frac{\partial \mathcal{L}}{\partial z(-i)} = -2\alpha \sum_{j=0}^{i} \left\{ W_{D\{p-i+j\}} \left(z(1+j) - \sum_{k=1}^{p} W_{D\{k\}} z(1+j-k) - \sum_{u=1}^{\#\,X\,comp.} W_{X\{u\}} x(u, 1+j) \right) \right\} \; with\; 0 \le i \le p-1$$

Eq. 38

The components of the gradient corresponding to the autoregressive coefficients in the state equation are as follows:

$$\forall t, \frac{\partial \mathcal{L}}{\partial W_{D\{i+1\}}} = -2\alpha z(t-i) \left(z(t) - \sum_{j=1}^{p} W_{D\{j\}} z(t-j) \right) + \gamma_W \frac{\partial \left\| W_{D\{i\}} \right\|_1}{\partial W_{D\{i\}}} \; with\; 0 \le i \le p-1$$

Eq. 39

$$\forall\, t, \frac{\partial \mathcal{L}}{\partial W_O} = -2\beta z(t)\big(y(t) - W_O z(t)\big) + \gamma_W \frac{\partial \|W_O\|_1}{\partial W_O}$$

Eq. 40

$$\forall\, t, \frac{\partial \mathcal{L}}{\partial W_X} = -2\beta \left(\sum_{u=1}^{\#\,X\,comp.} x_{\{u\}}(t) \right) \left(z(t) - \sum_{i=1}^{p} \sum_{j=1}^{p} W_{D\{i\}} z(t-i) - \sum_{u=1}^{\#\,X\,comp.} x_{\{u\}}(t) \right) + \gamma_W \frac{\partial \|W_X\|_1}{\partial W_X}$$

Eq.41

where $\|.\|_1$ corresponds to the L1-norm.

Finally, for the latent variable sequence Z, the gradient components are given by:

$$\frac{\partial \mathcal{L}}{\partial z(t)} = 2\gamma_Z\big(2z(t) - z(t-1) - z(t+1)\big) - 2\beta W_O\big(y(t) - W_O z(t)\big)$$
$$- \alpha\left[\left(z(t) - \sum_{i=1}^{p} W_{D\{i\}} z(t-i) - \sum_{u=1}^{\#\,X\,comp.} x_{\{u\}}(t)\right)\right.$$
$$\left. + \sum_{j=1}^{p} W_{D\{j\}}\left(z(t+j) - \sum_{k=1}^{p} W_{D\{k\}} z(t+j-k)\right)\right]$$
$$with\ 1 \le t \le T$$

Eq.42

We make use of a numerical approximation to the gradient components of $\mathcal{L}$ corresponding to the Z sequence as well as to the parameters W.

$$\frac{\partial \mathcal{L}}{\partial Z} = \frac{\mathcal{L}(Y, Z+\varepsilon; X, W) - \mathcal{L}(Y, Z-\varepsilon; X, W)}{2\varepsilon}$$
$$\frac{\partial \mathcal{L}}{\partial W} = \frac{\mathcal{L}(Y, Z; X, W+\varepsilon) - \mathcal{L}(Y, Z; X, W-\varepsilon)}{2\varepsilon}\ with\ \varepsilon = 10^{-6}$$

Eq. 43

Energy Systems Division
9700 South Cass Avenue, Bldg. 362
Argonne, IL 60439-4815

www.anl.gov

Argonne National Laboratory is a U.S. Department of Energy
laboratory managed by UChicago Argonne, LLC